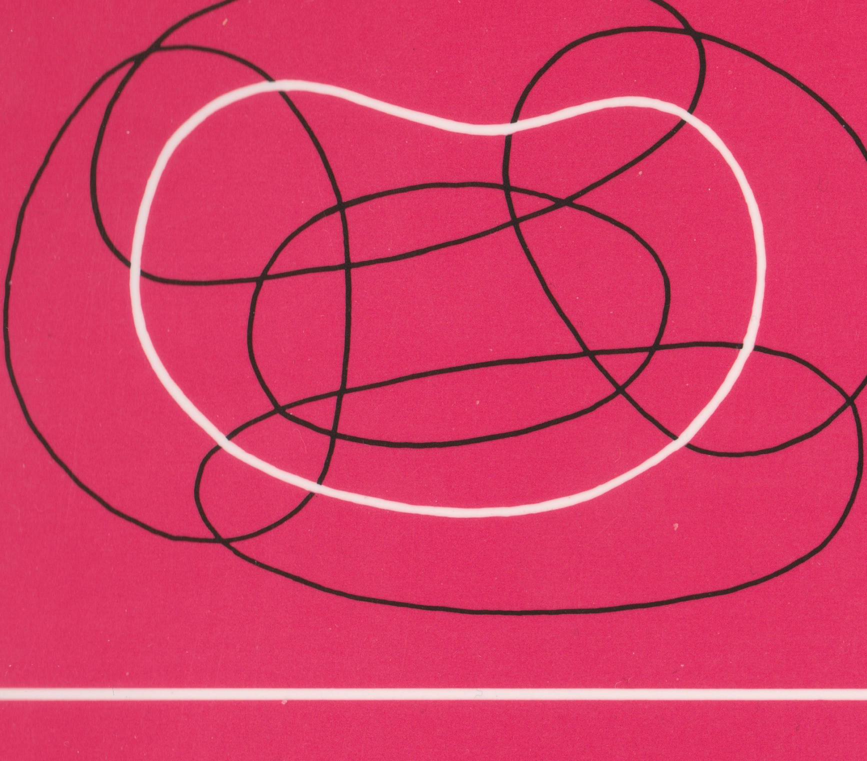
W. A. SUTHERLAND

Introduction to Metric and Topological Spaces



OXFORD SCIENCE PUBLICATIONS

Introduction to metric and topological spaces

W. A. SUTHERLAND

Introduction to metric and topological spaces

CLARENDON PRESS · OXFORD

Oxford University Press, Great Clarendon Street, Oxford OX2 6DP

Oxford New York
Athens Auckland Bangkok Bogota Bombay Buenos Aires
Calcutta Cape Town Dar es Salaam Delhi Florence Hong Kong Istanbul
Karachi Kuala Lumpur Madras Madrid Melbourne Mexico City
Nairobi Paris Singapore Taipei Tokyo Toronto Warsaw

and associated companies in
Berlin Ibadan

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press Inc., New York

© Oxford University Press, 1975

Reprinted 1981 (with corrections), 1983, 1985, 1986, 1987, 1988, 1989, 1991, 1993 (twice), 1995 (with corrections), 1998

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press. Within the UK, exceptions are allowed in respect of any fair dealing for the purpose of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms and in other countries should be sent to the Rights Department, Oxford University Press, at the address above.

This book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

ISBN 0 19 853161 3

Printed and bound in Great Britain by
Biddles Ltd, Guildford and King's Lynn

Preface

ONE of the ways in which topology has influenced other branches of mathematics in the past few decades is by putting the study of continuity and convergence into a general setting. This book introduces metric and topological spaces by describing some of that influence. The aim is to move gradually from familiar real analysis to abstract topological spaces; the main topics in the abstract setting are related back to familiar ground as far as possible. Apart from the language of metric and topological spaces, the topics discussed are compactness, connectedness, and completeness. These form part of the central core of general topology which is now used in several branches of mathematics. The emphasis is on *introduction;* the book is not comprehensive even within this central core, and algebraic and geometric topology are not mentioned at all. Since the approach is via analysis, it is hoped to add to the reader's insight on some basic theorems there (for example, it can be helpful to some students to see the Heine–Borel theorem and its implications for continuous functions placed in a more general context).

The stage at which a student of mathematics should see this process of generalization, and the degree of generality he should see, are both controversial. I have tried to write a book which students can read quite soon after they have had a course on analysis of real-valued functions of one real variable, not necessarily including uniform convergence.

The first chapter reviews real numbers, sequences, and continuity for real-valued functions of one real variable. Most readers will find nothing new there, but we shall continually refer back to it. With continuity as the motivating concept, the setting is generalized to metric spaces in Chapter 2 and to topological spaces in Chapter 3. The pay-off begins in Chapter 5 with the study of compactness, and continues in later chapters on connectedness and completeness. In order to introduce uniform convergence, Chapter 8 reverts to the traditional approach for real-valued functions of a real variable before interpreting this as convergence in the sup metric.

Most of the methods of presentation used are the common property of many mathematicians, but I wish to acknowledge that the way of introducing compactness is influenced by Hewitt (1960). It is also a pleasure to acknowledge the influence of many teachers, colleagues, and ex-students on this book, and to thank Peter Strain of the Open University for helpful comments and the staff of the Clarendon Press for their encouragement during the writing.

W. A. S.

Oxford, 1974

Preface to reprinted edition

I am grateful to all who have pointed out errors in the first printing (even to those who mentioned that the proof of Corollary 1.1.7 purported to establish the existence of a *positive* rational number between any two real numbers). In particular it is a pleasure to thank Roy Dyckhoff, Ioan James and Richard Woolfson for valuable comments and corrections.

W. A. S.

Oxford, 1981

Contents

INTRODUCTION ix

NOTATION AND TERMINOLOGY xi

1. REVIEW OF SOME REAL ANALYSIS 1
   - 1.1. Real numbers 1
   - 1.2. Real sequences 5
   - 1.3. Limits of functions 11
   - 1.4. Continuity 12

2. CONTINUITY GENERALIZED: METRIC SPACES 19
   - 2.1. Motivation 19
   - 2.2. Examples 22
   - 2.3. Open sets in metric spaces 33
   - 2.4. Equivalent metrics 38
   - 2.5. Continuity 42

3. CONTINUITY GENERALIZED: TOPOLOGICAL SPACES 45
   - 3.1. Topological spaces 45
   - 3.2. Bases 49
   - 3.3. Sub-bases and weak topologies 50
   - 3.4. Subspaces 51
   - 3.5. Products 53
   - 3.6. Homeomorphisms 56
   - 3.7. Definitions 59
   - 3.8. Quotient spaces 65

4. THE HAUSDORFF CONDITION 72
   - 4.1. Motivation 72
   - 4.2. Separation axioms 73

5. COMPACT SPACES 75
   - 5.1. Motivation 75
   - 5.2. Definition of compactness 77
   - 5.3. Compactness of $[a, b]$ 79
   - 5.4. Properties of compact spaces 82
   - 5.5. Continuous maps on compact spaces 83
   - 5.6. Compactness and constructions 84
   - 5.7. Compact subspaces of $\mathbf{R}^n$ 86
   - 5.8. Compactness and uniform continuity 87
   - 5.9. An inverse function theorem 88

6. CONNECTED SPACES 93
   - 6.1. Introduction 93
   - 6.2. Connectedness 93
   - 6.3. Path-connectedness 99
   - 6.4. Comparison of definitions 102
   - 6.5. Components 103

7. COMPACTNESS AGAIN: CONVERGENCE IN METRIC SPACES 108
   - 7.1. Introduction 108
   - 7.2. Sequential compactness 108

8. UNIFORM CONVERGENCE 114
   - 8.1. Introduction 114
   - 8.2. Definition and examples 114
   - 8.3. Cauchy's criterion 119
   - 8.4. Uniform limits of sequences 120
   - 8.5. Generalizations 121

9. COMPLETE METRIC SPACES 123
   - 9.1. Introduction 123
   - 9.2. Definition and examples 123
   - 9.3. Fixed point theorems 129
   - 9.4. The contraction mapping theorem 130
   - 9.5. Cantor's and Baire's theorems 133

10. CRITERIA FOR COMPACTNESS IN METRIC SPACES 139
   - 10.1. A general criterion 139
   - 10.2. Arzelà–Ascoli theorem 141

11. APPENDIX 149
   - 11.1. Real numbers 149
   - 11.2. Completion of metric spaces 155

12. GUIDE TO EXERCISES 160

BIBLIOGRAPHY 177

INDEX 179

Introduction

In this book we are going to generalize some theorems about convergence and continuity which are probably familiar to the reader in the case of sequences of real numbers and real-valued functions of one real variable. An example of the kind of result we shall be aiming to generalize is the following: *if f is a real-valued function which is defined and continuous on the closed interval $[a, b]$ in the real line, then f is bounded on $[a, b]$, i.e., there exists a real number K such that $|f(x)| \leqslant K$ for all x in $[a, b]$.* Several such theorems about real-valued functions of a real variable are true and useful in a more general framework, after suitable minor changes of wording. For example if we suppose that f is a real-valued function of two real variables which is defined and continuous in a rectangle $[a, b] \times [c, d]$, then it is still true that f is bounded on this rectangle. Once we have seen that the result generalizes from one to two real variables, it is natural to suspect that it is true for any finite number of real variables, and then to go a step further by asking just how general a situation can the theorem be formulated for, and how generally is it true? These questions lead us to metric and topological spaces.

Before going on to study such questions, it is fair to ask: what is the point of generalization? One answer is simply that it saves time, or at least avoids tedious repetition. For example, if we can show by a single proof that a certain result holds for functions of n real variables, where n is any positive integer, this is better than proving it separately for one real variable, two real variables, three real variables, etc. In the same vein, generalization often gives a unified mental grasp of several results which otherwise might just seem vaguely similar, and in addition to the satisfaction involved, this more efficient organization of material helps some people's understanding. Another gain is that generalization often illuminates the proof of a theorem, because to see how generally a given result can be proved, one has to notice exactly what properties or hypotheses are used at each stage in the proof. A motive which is perhaps more sophisticated is the urge to prove any given result in the appropriate context; what this means is partly a matter of taste, and it is not easy to explain before it has been done.

Against these motives, we should be aware of some dangers in generalization. Most mathematicians would agree that it can be carried to an excessive and useless extent. Just when this stage is reached is a matter of controversy, but the potential reader is warned that some mathematicians would say 'Enough, no more (at least as far as analysis is concerned)' when we get into metric spaces. Also, there is an initial barrier of unfamiliarity to be overcome in moving to a more general framework, with its new language; the extent to which the pay-off is worthwhile is likely to vary from one student to another.

As a consequence of introducing abstractions gradually, the theorem density is low. The title of theorem is reserved for substantial results, which have significance in a broad range of mathematics.

Some paragraphs and exercises are marked † to denote that they require some knowledge of abstract algebra, and others are marked * to denote that they are tentatively thought to be more severe than the rest. There are hints at the end of the book for all such exercises and some others.

At several points in the text, suggestions are made about helpful diagrams which can be drawn. Readers are strongly urged to draw their own diagrams wherever possible, even when no specific suggestions are made.

A previous course in real analysis is regarded as a prerequisite for reading this book. What this means is an introduction (including rigorous proofs) to continuity, differential (and preferably also integral) calculus for real-valued functions of one real variable, and to convergence of real number sequences and series. This material is contained for example in Burkill (1970) or Spivak (1973) (names followed by dates in parentheses refer to the bibliography at the end of the book). Chapter 1 is intended only as a review of some of the real analysis to which the rest of the book most frequently refers. It does not include special functions and facts from calculus which we sometimes use for illustration. The experience of abstraction gained from a previous course in, say, linear algebra would help the reader in a general way to follow the abstraction of metric and topological spaces. However, the student is likely to be the best judge of whether she/he is ready, or wants, to read this book.

Notation and terminology

WE use the logical symbols $\Rightarrow$ and $\Leftrightarrow$ or iff, meaning *implies* and *if and only if*. Throughout, we use the language of sets and maps. Since most introductions to algebra and analysis contain a survey of this language, we merely list notation.

If an object a belongs to a set A we write $a \in A$, and if not we write $a \notin A$. If A is a subset of B (perhaps equal to B) we write $A \subset B$ or $B \supset A$. The subset of elements of A possessing some property P is written $\{a \in A : P(a)\}$. A finite set is sometimes specified by listing its elements, say $\{a_1, a_2, \ldots, a_n\}$. Intersection and union of sets are denoted by $\bigcap$ and $\bigcup$. The empty set is written $\varnothing$. Given two sets A and B, the set $\{b \in B : b \notin A\}$ is written $B - A$. Thus in particular if $A \subset B$, the complement of A in B is $B - A$. If for each i in some set I we are given a subset A_i of a set S, then we denote by $\bigcup_{i \in I} A_i$, $\bigcap_{i \in I} A_i$ (or just $\bigcup_I A_i$, $\bigcap_I A_i$) the union and intersection over all i in I; in this situation, I is called an *indexing* set. We use De Morgan's laws, which in the above notation assert

$$S - \bigcup_I A_i = \bigcap_I (S - A_i), \qquad S - \bigcap_I A_i = \bigcup_I (S - A_i).$$

The Cartesian product $A \times B$ of sets A, B is the set of all ordered pairs (a, b) where $a \in A$, $b \in B$. This generalizes easily to the product of any finite number of sets; in particular we use A^n to denote the set of ordered n-tuples of elements from A. We shall not consider infinite products of sets. Occasionally we use 2^A to denote the set of all subsets of a set A.

A map or function (the terms are used interchangably) between sets A, B is written $f : A \rightarrow B$. We call A the *domain* of f, and we avoid calling B anything. We think of f as assigning to each a in A an element $f(a)$ in B, although logically it is preferable to define a map as a pair of sets A, B together with a certain type of subset of $A \times B$ (intuitively the graph of f). Persisting with our way of thinking of f, we define the *graph of* f to be $G_f = \{(a, b) \in A \times B : f(a) = b\}$. For any subset $C \subset A$, the (*direct*) *image* $f(C)$ of C under f is $\{b \in B : b = f(c)$ for some c in $C\}$ and for any subset $D\subset B$, the *inverse image* $f^{-1}(D)$ of D under f is $\{a\in A: f(a)\in D\}$. We use the formulae

$$f\Big(\bigcap_I C_i\Big)\subset\bigcap_I f(C_i),\qquad f\Big(\bigcup_I C_i\Big)=\bigcup_I f(C_i),$$

$$f^{-1}\Big(\bigcap_I D_i\Big)=\bigcap_I f^{-1}(D_i),\qquad f^{-1}\Big(\bigcup_I D_i\Big)=\bigcup_I f^{-1}(D_i),$$

$$f^{-1}(B-D_i)=A-f^{-1}(D_i),$$

where $f: A\to B$ is a map and $C_i\subset A$, $D_i\subset B$ for every i in I. Note that equality does not necessarily hold in the first of these. We call f *injective* if $f(a)=f(a')\Rightarrow a=a'$, since the term one–one is a little ambiguous. We should therefore call $f: A\to B$ surjective if $f(A)=B$, but instead we call such an f *onto*. If f has both these properties we call it a *one–one correspondence*. In the special case when a map $f: A\to B$ is injective we may define an inverse function $f^{-1}: f(A)\to A$. When f is injective and $C\subset f(A)$ there is therefore a second meaning of $f^{-1}(C)$. Fortunately this determines the same subset of A as the previous meaning, but we emphasize that the inverse image set $f^{-1}(C)$ is defined for any $C\subset B$, whether f is injective or not, whereas the inverse map f^{-1} is defined only if f is injective, and even when f is injective the direct image $f^{-1}(C)$ of C under f^{-1} is defined only if $C\subset f(A)$.

Given a map $f: A\to B$ and a subset $C\subset A$ we write $f\mid C: C\to B$ for the *restriction* of f to C, defined by $(f\mid C)(c)=f(c)$ for all c in C. In this situation we also refer to f as an *extension* of $f\mid C$ to A. Given maps $f: A\to B$, $g: B\to C$, the *composition* $g\circ f: A\to C$ is defined by $g\circ f(a)=g(f(a))$ for any a in A. In particular if $i: D\to A$ denotes the inclusion map of a subset D in A, and $f: A\to B$ is any map, then $f\circ i=f\mid D$. Special maps used include the identity map $f: A\to A$ of any set A, given by $f(a)=a$ for all a in A, and constant maps $f: A\to B$, given by $f(x)=b$ for every x in A and some fixed b in B.

We shall occasionally assume that the terms *equivalence relation* and *countable set* are understood.

We use **N**, **Z**, **Q**, **R**, **C** to denote the sets of natural numbers (or positive integers), integers, rational numbers, real numbers, and complex numbers respectively. We often refer to **R** as the *real line*, and we call the following subsets of **R** *intervals:*

(i) $[a, b]=\{x\in\mathbf{R} : a\leqslant x\leqslant b\}$,
(ii) $(a, b)=\{x\in\mathbf{R} : a< x< b\}$,
(iii) $(a, b]=\{x\in\mathbf{R} : a< x\leqslant b\}$,
(iv) $[a, b)=\{x\in\mathbf{R} : a\leqslant x< b\}$,
(v) $(-\infty, b]=\{x\in\mathbf{R} : x\leqslant b\}$,
(vi) $(-\infty, b)=\{x\in\mathbf{R} : x< b\}$,
(vii) $[a, \infty)=\{x\in\mathbf{R} : x\geqslant a\}$,
(viii) $(a, \infty)=\{x\in\mathbf{R} : x> a\}$,
(ix) $(-\infty, \infty)=\mathbf{R}$.

The intervals in (i), (v), (vii) (and (ix)) are called *closed* intervals; those in (ii), (vi), (viii) (and (ix)) are called *open* intervals, and the others are called *half-open* intervals. When we refer to an interval of types (i)–(iv), it is always to be understood that $b>a$, except for type (i), when on stated occasions we also allow $b=a$. We shall try to avoid the occasional risk of confusing an interval (a, b) in **R** with a point (a, b) in $\mathbf{R}^2$ by stating which of these is meant when there might be any doubt.

1

Review of some real analysis

1.1 Real numbers

Two popular ways of thinking about the real number system are:

(1) geometrically, as corresponding to all the points on a straight line;

(2) in terms of decimal expansions, where if a number is irrational we think of longer and longer decimal expansions approximating it more and more closely.

Neither of these intuitive ideas is precise enough for our purposes, although each leads to a way of constructing the real numbers from the rational numbers. The second of these ways is described in the appendix. As the reader probably knows, one approach to real numbers is to use a set of axioms. This leaves aside the questions of whether there is any system satisfying the axioms and to what extent such a system is unique. Also, it can hardly be said to explain what the real numbers are to anyone who does not already know in an intuitive sense. But it has the merit of providing quickly a point of departure for analysis.

Many introductions to analysis contain a list of axioms for the real numbers (see, for example, Spivak (1973) or Chapter 1 of Apostol (1957)). A large number of these axioms may be summed up technically by saying that the real numbers form an ordered field. A less precise description is that addition, subtraction, multiplication, and division of real numbers all work in the way that we expect them to, and that the same is true of the way in which inequalities $x<y$ work and interact with addition and multiplication. We shall not review these axioms, but concentrate solely on the so-called completeness axiom. The reasons for this strange behaviour are, first, that this is the axiom which distinguishes the real numbers from the rational numbers (and in a sense analysis from algebra), and secondly that our intuition is unlikely

to let us down on properties deducible from the ordered field axioms, whereas arguments using the completeness axiom tend to be more subtle.

In discussing the completeness axiom we shall proceed as if the real number system were already 'there', and our job is to describe one property of it. (In a strictly axiomatic approach we ought to explain what the completeness axiom says using only the ordered field axioms, and then define the real numbers to be any system satisfying the completeness axiom as well as the ordered field axioms.)

We need some terminology. Let S be a non-empty set of real numbers. An upper bound for S is a number x such that $y \leqslant x$ for all y in S. If an upper bound for S exists we say that S is bounded above. Lower bounds are defined similarly.

EXAMPLES 1.1.1. (a) The set $\mathbf{R}$ of all real numbers has no upper or lower bound.

(b) The set $\mathbf{R}_-$ of all strictly negative real numbers has no lower bound, but for example 0 is an upper bound.

(c) The half-open interval $(0, 1]$ is bounded above and below.

If S has an upper bound u, then S has many upper bounds, since any x satisfying $x \geqslant u$ is another upper bound. This gives the next definition some point.

DEFINITION 1.1.2. Given a non-empty subset S of $\mathbf{R}$ which is bounded above, we call u a *least upper bound for S* if

(a) u is an upper bound for S,
(b) $x \geqslant u$ for any upper bound x of S.

EXAMPLE 1.1.3. In Example 1.1.1(b), 0 is a least upper bound for $\mathbf{R}_-$. For 0 is an upper bound, and it is a *least* upper bound because no strictly negative number x can be an upper bound for $\mathbf{R}_-$, since $\frac{1}{2}x > x$ and $\frac{1}{2}x \in \mathbf{R}_-$. Examples 1.1.1(c) and (b) show that a least upper bound of a set S may or may not be in S.

It follows from Definition 1.1.2 that least upper bounds are unique when they exist. For if u, u' are both least upper bounds for a set S, then since u' is an upper bound it follows that $u \leqslant u'$ by leastness of u (by leastness we mean property 1.1.2(b)). Similarly $u' \leqslant u$, so $u = u'$. Greatest lower bounds are defined similarly. We can now state one form of the completeness axiom for $\mathbf{R}$.

AXIOM 1.1.4. *Any non-empty subset of* **R** *which is bounded above has a least upper bound.*

This axiom is quite subtle, and it is difficult to grasp its full significance until it has been used several times. It corresponds to the intuitive idea that there are no gaps in the real numbers, thought of as the points on a straight line; but the transition from the intuitive idea to the formal statement is not immediately obvious. For some sets of real numbers, such as Examples 1.1.1(b) and (c), it is 'obvious' that a least upper bound exists (strictly speaking, this means that it follows from the ordered field axioms). As a first example of the kind of set to which we really want to apply the completeness axiom, and of its relation to the existence of irrational numbers, let $S=\{x\in\mathbf{Q}:x^2<2\}$. Then intuitively the least upper bound of S is the irrational $\sqrt{2}$ (we write $\sqrt{2}$ for the positive square root of 2). Formally, we need to prove the following assertions:

(a) if u is a least upper bound for S then $u^2=2$,
(b) there is no rational number u such that $u^2=2$.

We shall prove (a) in Example 1.1.8 below. The reader has probably seen a proof of (b); it is included in Exercise 1.5.5 for which hints are given.

For any non-empty subset S of **R** which is bounded above, we call the unique least upper bound *sup* S (sup is short for supremum. Other notation sometimes used is l.u.b. S). Although the completeness axiom was stated in terms of sets bounded above, it is equivalent to the corresponding property for sets bounded below. The next proposition formally states half of this equivalence.

PROPOSITION 1.1.5. *If a non-empty subset S of* **R** *is bounded below then it has a greatest lower bound.*

Proof. Let $T=\{x\in\mathbf{R}:-x\in S\}$. The idea of the proof is simply that l is a lower bound for S iff $-l$ is an upper bound for T. The details are left as Exercise 1.5.6.

Just as in the case of least upper bounds, a non-empty subset S of **R** which is bounded below has a unique greatest lower bound called *inf* S (short for infimum) or g.l.b.S.

We next give two applications of the completeness axiom. The proofs use several 'obvious' facts about **R** which in a strictly axiomatic approach would be deduced from the ordered field axioms. We leave the reader to spot these.

PROPOSITION 1.1.6. *The set* **N** *of natural numbers is not bounded above.*

Proof. Suppose that **N** is bounded above. Then by the completeness axiom there is a real number $u = \sup \mathbf{N}$. For any n in **N**, $n+1$ is also in **N**, so $n+1 \leqslant u$. But then $n \leqslant u-1$. Since this is true for all n in **N**, $u-1$ is an upper bound for **N**, contradicting the leastness of u. This contradiction shows that **N** cannot be bounded above.

The property expressed in Proposition 1.1.6, or some equivalent statement, is called Archimedes' axiom (it is an axiom in some axiom systems for **R**, although not in ours).

COROLLARY 1.1.7. *Between any two distinct real numbers x, y there is a rational number.*

Proof. Suppose $x<y$, and first suppose $x \geqslant 0$. Since $y-x>0$, by 1.1.6 there is an n in **N** satisfying $n>1/(y-x)$ so $1/n<y-x$. Let $M=\{m \in \mathbf{N} : m/n > x\}$. By 1.1.6, $M \neq \varnothing$, otherwise nx would be an upper bound for **N**. Hence, since $M \subset \mathbf{N}$, M contains a least integer v. This means $v/n>x$, $(v-1)/n \leqslant x$. Hence $x<v/n \leqslant x + 1/n < x+(y-x)=y$, and v/n is a suitable rational number. Now suppose $x<0$. If $y>0$ then 0 is a rational such that $x<0<y$. If $y \leqslant 0$, the previous case supplies a rational r such that $-y<r<-x$, so $x<-r<y$.

REMARK. Between any two distinct real numbers there is also an irrational number (see Exercise 1.5.7).

We are now in a position to prove the existence of $\sqrt{2}$.

EXAMPLE 1.1.8. There exists a real number u such that $u^2=2$.

Proof. Let $S=\{x \in \mathbf{Q} : x^2<2\}$ (we could equally well use $\{x \in \mathbf{R} : x^2<2\}$). Then S is non-empty ($1 \in S$). To prove that S is bounded above, there is no need to be fastidious—the existence of some upper bound, however much larger than is strictly necessary, will do. For example, 10 is an upper bound, since if $y>10$ then $y^2>100>2$, and y is not in S. Hence $x \in S \Rightarrow x \leqslant 10$. Now by the completeness axiom, $u = \sup S$ exists. In fact $u \geqslant 1$, since $1 \in S$. We shall prove that $u^2=2$, by showing that each of $u^2>2$, $u^2<2$ leads to a contradiction.

First suppose that $u^2>2$. Then $(u^2-2)/2u>0$, so by Proposition 1.1.6 there exists an n in $\mathbf{N}$ satisfying

$$0<1/n<(u^2-2)/2u.$$

Then

$$(u-1/n)^2=u^2-2u/n+1/n^2>u^2-2u/n>u^2-(u^2-2)=2,$$

so $x\in S\Rightarrow x^2<2<(u-1/n)^2\Rightarrow x<u-1/n$, contradicting the leastness of u.

Secondly suppose that $u^2<2$. Choose an integer n such that $0<1/n\leqslant(2-u^2)/4u$ and $1/n<2u$. Then

$$\begin{aligned}(u+1/n)^2=u^2+2u/n+1/n^2&<u^2+2u/n+2u/n \quad (\text{since } 1/n<2u)\\ &\leqslant u^2+2-u^2 \quad (\text{since } 2-u^2\geqslant 4u/n)\\ &=2.\end{aligned}$$

Hence $u+1/n\in S$, contradicting the fact that u is an upper bound for S. This completes the proof that $u^2=2$.

By similar proofs we could establish the existence of other square roots, cube roots, etc. However, the kind of work needed above can later be done much more efficiently, although the proofs are still based on the completeness axiom.

The completeness axiom can be stated in various other forms. Two of these are mentioned in §1.2 below.

We conclude this brief review of real numbers by recalling two useful inequalities, whose proofs are immediate.

PROPOSITION 1.1.9. *$|x+y|\leqslant|x|+|y|$ for any x, y in* $\mathbf{R}$.

COROLLARY 1.1.10. *$|x-y|\geqslant\big||x|-|y|\big|$ for any x, y in* $\mathbf{R}$.

1.2 Real sequences

Formally an infinite sequence of real numbers is a map $s:\mathbf{N}\to\mathbf{R}$. This definition is useful for discussing topics such as subsequences and rearrangements without being vague. In practice, however, given such a map s we denote $s(n)$ by s_n and think of the sequence in the traditional way as an infinite ordered string of numbers, using the notation (s_n) or $s_1, s_2, s_2, \ldots$ instead of s for the whole sequence.

It is important to distinguish between a sequence (s_n) and the set of its members, $\{s_n: n\in\mathbf{N}\}$. The latter can easily be finite. For example if (s_n) is 1, 0, 1, 0, …, then the set of its members is $\{0, 1\}$. Formally, this is a matter of distinguishing between a map $s:\mathbf{N}\to\mathbf{R}$ and its image set $s(\mathbf{N})$.

Sequences can arise for example in solving algebraic or differential equations. On the practical side, s_n might be the answer at the nth stage in some method of successive approximations for finding a root of an equation. On the theoretical side, convergent sequences might be used to prove the existence of solutions to equations. The only difference between theory and practice here is that in practice one is interested in how quickly the sequence gives a good approximation to the answer. In applications, the members of the sequence might be vectors or functions instead of real numbers. In this section we review real number sequences, with the emphasis on definitions and theoretical results.

EXAMPLES 1.2.1. (a) $\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots,$
(b) $1, -\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}, \ldots,$
(c) $\frac{1}{2}, 1, -\frac{1}{2}, -1, \frac{1}{4}, \frac{1}{2}, \ldots,$
(d) $1, 2, 3, \ldots,$
(e) $1, 0, 1, 0, \ldots.$

In these examples, there is a simple formula for s_n in terms of n, which the reader will spot. This is convenient for illustrating the basic theory of sequences, but in practice a sequence might be generated by an iterative process, or by the results of a probabilistic experiment repeated more and more often, or by some other means, and in such cases there may not be any simple formula for s_n in terms of n.

EXAMPLES 1.2.1. (f) $s_1=1$, $s_2=0$, $s_n=\frac{1}{2}(s_{n-2}+s_{n-1})$ for $n>2$,
(g) s_n is the answer at the nth stage in some specified algorithm for approximating $\sqrt{2}$.

In Examples (a), (b), (c) the sequence seems intuitively to be heading towards a definite number, whether steadily, by bracketing the target, or irregularly, whereas in Examples (d), (e) this is not the case. The mathematical term for ‘heading towards’ is ‘converging’ and the precise definition is as follows.

DEFINITION 1.2.2 The sequence (s_n) *converges to* (the real number) l, if given any (real number) $\varepsilon>0$, there exists (an integer) N such that $|s_n-l|<\varepsilon$ for all $n\geqslant N$.

This is usually shortened by omitting the parts in parentheses.

Other ways of writing '(s_n) converges to l' are '$s_n \to l$ as $n \to \infty$' and '$\lim_{n\to\infty} s_n = l$'. Here are two ways of thinking about the definition.

(1) (s_n) converges to l if, given any required degree of accuracy, then by going far enough along the sequence we can be sure that the terms beyond that stage all approximate l to within the required degree of accuracy.

(2) Let us take coordinate axes in the plane and mark the points with coordinates (n, s_n). Let us also draw a horizontal line L at height l. Then (s_n) converges to l if, given any horizontal band of positive width centred on L, there exists a vertical line such that all marked points to the right of that vertical lie within the prescribed horizontal band. Figure 1.1 is the kind of picture this suggests. The sequence promises to stay out of the shaded territory.

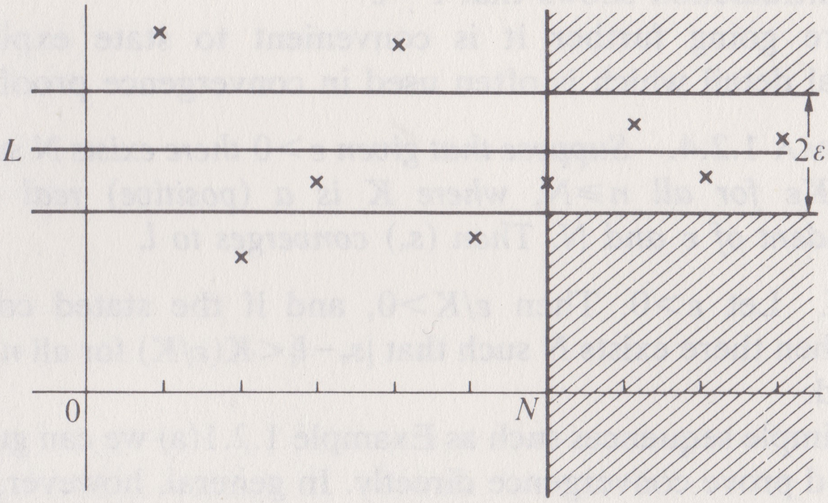


FIG. 1.1

The reader is probably aware of two points which are easy to get wrong when one is first trying to wield the formal definition, even although the intuitive idea is understood. First, the order in which ε, N occur is crucial: given any $\varepsilon > 0$ *first*, there must *then* be an N, which for emphasis we might label N_ε, such that ... etc. Secondly, to prove convergence it is not enough to show that given $\varepsilon > 0$, there exists an N such that $|s_n - l| < \varepsilon$ for *some* $n \geqslant N$: for example this would be true of the sequence 1, 0, 1, 0, ..., with $l = 0$, any $\varepsilon > 0$, and $N = 1$, yet the sequence does not converge.

The first deduction from the formal definition is an obvious part of the intuitive idea of convergence.

PROPOSITION 1.2.3. *A convergent sequence has a unique limit.*

Proof. Suppose that (s_n) converges to l and also to l', where $l' \neq l$. Put $\varepsilon = \frac{1}{2}|l-l'|$. Since (s_n) converges to l, there exists N_1 such that $|s_n - l| < \varepsilon$ for all $n \geqslant N_1$. Similarly there exists N_2 such that $|s_n - l'| < \varepsilon$ for all $n \geqslant N_2$. Put $N = \text{Max}\{N_1, N_2\}$. Then

$$|l-l'| = |l - s_N + s_N - l'| \leqslant |l - s_N| + |s_N - l'| < 2\varepsilon = |l-l'|.$$

This contradiction shows that $l' = l$.

Before going further it is convenient to state explicitly a technical detail which is often used in convergence proofs.

LEMMA 1.2.4. *Suppose that given $\varepsilon > 0$ there exists N such that $|s_n - l| < K\varepsilon$ for all $n \geqslant N$, where K is a (positive) real number independent of ε and N. Then (s_n) converges to l.*

Proof. Let $\varepsilon > 0$. Then $\varepsilon/K > 0$, and if the stated condition holds, then there exists N such that $|s_n - l| < K(\varepsilon/K)$ for all $n \geqslant N$, as required.

For simple sequences such as Example 1.2.1(a) we can guess the limit and prove convergence directly. In general, however, it may be hard to guess the limit, and more importantly there may be no more convenient way to name a real number than as the limit of a given sequence. The meaning of the latter statement brings us back to the definition of real numbers, and is probably best explained by the approach in the appendix, but as an example consider

$$s_n = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!}.$$

The reader may be able to think of a way to define e other than as the limit of (s_n), but it will also directly or indirectly involve taking the limit of this or some other sequence (such as $t_n = (1+1/n)^n$).

We shall consider two theorems which provide ways of proving convergence without using a known value of the limit. As the above discussion indicates, both will depend heavily on the completeness axiom for $\mathbf{R}$, and in fact each leads to another way of stating that axiom.

DEFINITION 1.2.5. A sequence (s_n) is *monotonic increasing* (*decreasing*) if $s_{n+1} \geqslant s_n$ $(s_{n+1} \leqslant s_n)$ for all n in $\mathbf{N}$. It is *monotonic* if it has either of these properties.

THEOREM 1.2.6. *Every bounded monotonic sequence converges.*

Proof. Suppose for example that (s_n) is monotonic increasing and bounded above. This means that the set S of its members is bounded above, so by the completeness axiom $l = \sup S$ exists. We shall prove that (s_n) converges to l. Let $\varepsilon > 0$. Then $l - \varepsilon$ is not an upper bound for S, by leastness of l, so $s_N > l - \varepsilon$ for some N. Since (s_n) is monotonic increasing, $s_n \geqslant s_N$ for all $n \geqslant N$. Also, $s_n \leqslant l$ for all n, since l is an upper bound for S. Hence for all $n \geqslant N$, $l \geqslant s_n \geqslant s_N > l - \varepsilon$, and in particular $|l - s_n| < \varepsilon$.

The proof for (s_n) monotonic decreasing is similar. We state without proof a result in the other direction, for which *convergence* should really be redefined as on p. 149.

THEOREM 1.2.7. *Suppose that in an ordered field, Archimedes' axiom holds and every monotonic increasing sequence which is bounded above converges. Then every non-empty subset which is bounded above has a least upper bound.*

As well as being useful on its own, Theorem 1.2.6 helps to prove the next convergence criterion. First we give a name to sequences in which the terms get closer and closer together as we get further out in the sequence.

DEFINITION 1.2.8. A sequence (s_n) is a *Cauchy* sequence if given $\varepsilon > 0$, there exists N such that $m, n \geqslant N$ (i.e. $m \geqslant N$, $n \geqslant N$) imply $|s_n - s_m| < \varepsilon$.

THEOREM 1.2.9. (Cauchy's convergence criterion.) *A sequence (s_n) of real numbers converges iff it is a Cauchy sequence.*

Proof. Suppose that (s_n) converges to l. Then given $\varepsilon > 0$, there exists N such that $|s_n - l| < \varepsilon$ for $n \geqslant N$, so for m, $n \geqslant N$ we have

$$|s_n - s_m| = |s_n - l + l - s_m| \leqslant |s_n - l| + |l - s_m| < 2\varepsilon.$$

Hence (cf. Lemma 1.2.4) (s_n) is a Cauchy sequence.

Suppose conversely that (s_n) is a Cauchy sequence. We show first that (s_n) is bounded. Take $\varepsilon = 1$, say, in the Cauchy condition. Thus there exists N such that m, $n \geqslant N$ imply $|s_n - s_m| < 1$, so for any $n \geqslant N$, $|s_n - s_N| < 1$, and hence $|s_n| < 1 + |s_N|$. Now $|s_n| \leqslant \operatorname{Max}\{|s_1|, |s_2|, \ldots, |s_{N-1}|, 1 + |s_N|\}$ for all n. We could equally well take $\varepsilon = 10^{10}$ or 10^{-10} in this part of the proof.

Next, in order to use Theorem 1.2.6, we manufacture a monotonic sequence out of (s_n) by the following subtle construction. For each m, let S_m denote the set of members of the sequence from the mth stage onwards, $S_m=\{s_n : n\geqslant m\}$. Since the whole set of members $S=S_1$ of the sequence is bounded, so is S_m. Hence by the completeness axiom, $\sup S_m$ exists. Let $t_m=\sup S_m$. Since $S_{m+1}\subset S_m$, it follows that $\sup S_{m+1}\leqslant \sup S_m$ (see Exercise 1.5.1). Thus (t_m) is monotonic decreasing. Also, $t_m\geqslant s_m$ by definition of t_m, and S is bounded below, hence so is (t_m). By Theorem 1.2.6, (t_m) converges, to l say.

Finally we shall prove by a 3ε-argument that (s_n) also converges to l. Given $\varepsilon>0$, there exists N_1 such that $|s_m-s_n|<\varepsilon$ for $m, n\geqslant N_1$ and N_2 such that $|l-t_m|<\varepsilon$ for $m\geqslant N_2$. Let $N=\operatorname{Max}\{N_1, N_2\}$. By definition of t_N, $t_N-\varepsilon$ is not an upper bound of $S_N=\{s_n : n\geqslant N\}$, so there exists $M\geqslant N$ such that $s_M>t_N-\varepsilon$; and $s_M\leqslant t_N$ since t_N is an upper bound for S_N. Now for any $n\geqslant N$,

$$|s_n-l|=|s_n-s_M+s_M-t_N+t_N-l|\leqslant|s_n-s_M|+|s_M-t_N|+|t_N-l|<3\varepsilon.$$

Hence (s_n) converges to l (using Lemma 1.2.4).

The proof of the following analogue of Theorem 1.2.7 is sketched in the appendix.

THEOREM 1.2.10. *Suppose that in an ordered field, Archimedes' axiom holds and every Cauchy sequence converges. Then every non-empty subset which is bounded above has a least upper bound.*

Before leaving sequences, we recall that their limits behave well under algebraic operations, in the following sense.

PROPOSITION 1.2.11. *Suppose that (s_n), (t_n) converge to l, m. Then*

(a) *(s_n+t_n) converges to $l+m$,*
(b) *(s_nt_n) converges to lm,*
(c) *$(1/t_n)$ converges to $1/m$ provided $m\neq 0$.*

To interpret (c) one shows that if $m\neq 0$ then t_n is non-zero (in fact $|t_n|$ is bounded away from zero) for all sufficiently large n.

We shall not review the way in which results about sequences translate into results about series, since series do not feature prominently in this book. Nevertheless it is worthwhile emphasizing the importance of practical experience with particular sequences and series, in understanding convergence. A few particular limits which we need are included for revision in Exercises 1.5. For many more examples, and for series we refer to Ferrar (1938).

1.3 Limits of functions

Limits of functions are used in the theoretical study of continuity, differentiation and integration, and in practical estimates of the behaviour of particular functions.

Suppose first for simplicity that we have a function $f: \mathbf{R} \to \mathbf{R}$. Let $a \in \mathbf{R}$.

DEFINITION 1.3.1. We say that $f(x)$ *tends to the limit* l as x tends to a $\left(\lim_{x\to a} f(x) = l\right)$ if given any (real number) $\varepsilon > 0$, there exists a (real number) $\delta > 0$ such that $|f(x) - l| < \varepsilon$ for all x satisfying $0 < |x - a| < \delta$.

This is similar to the definition of convergence of a sequence (s_n), but instead of looking at s_n for large values of n, we look at $f(x)$ for x close to, but not equal to, a. The value $f(a)$ is irrelevant to the existence of $\lim_{x\to a} f(x)$, and the limit, if it exists, may or may not equal $f(a)$.

EXAMPLE 1.3.2. Let $f: \mathbf{R} \to \mathbf{R}$ be given by

$$f(x) = x, \qquad x \neq 0, \qquad f(0) = 1.$$

Then $\lim_{x\to 0} f(x) = 0$. For given $\varepsilon > 0$, put $\delta = \varepsilon$. If $0 < |x - 0| < \delta$, then $|f(x) - 0| = |x| < \varepsilon$, as required.

To emphasize further that $f(a)$ is irrelevant to the existence of $\lim_{x\to a} f(x)$, we note that $f(a)$ need not even be defined, for Definition 1.3.1 to make sense—it is enough to assume that f is defined on $A \subset \mathbf{R}$ where, for some $b > 0$, A contains all x satisfying $0 < |x - a| < b$. Even this is more restrictive than is necessary for a sensible definition: the domain A need only contain numbers arbitrarily close to a. We shall not study this general case, but we note two specially useful ways of relaxing Definition 1.3.1. Suppose first that the domain A of f contains the open interval (a, b) for some $b > a$.

DEFINITION 1.3.3. The *right-hand limit* $\lim_{x\to a+} f(x) = l$ if given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - l| < \varepsilon$ for all x in $(a, a + \delta)$. (Note that δ may be chosen small enough so that $(a, a + \delta) \subset (a, b)$, and $f(x)$ is therefore defined for all x in $(a, a+\delta)$.) Left-hand limits are defined similarly.

Here are two examples much used in illustrating theoretical points.

EXAMPLES 1.3.4. Let f, $g:\mathbf{R}-\{0\}\to\mathbf{R}$ be given by $f(x)=x\sin 1/x$, $g(x)=\sin 1/x$. Then $\lim\limits_{x\to 0} f(x)=0$, and $\lim\limits_{x\to 0} g(x)$ does not exist. The proofs are left as Exercise 1.5.16.

Results about limits of functions may be proved by analogy with the proofs in § 1.2, or we may deduce them from § 1.2 using a conversion lemma.

LEMMA 1.3.5. *The following are equivalent:*

(i) $\lim\limits_{x\to a} f(x)=l$,

(ii) *if* (x_n) *is any sequence such that* (x_n) *converges to a, but for each n,* $x_n\neq a$, *then* $(f(x_n))$ *converges to l.*

We omit the proof. One may also prove analogues of Theorem 1.2.9 and Proposition 1.2.11 for limits of functions, and for left- and right-hand limits. The similarity between §§ 1.2 and 1.3 suggests that limits of sequences and limits of functions may both be special cases of a more general kind of limit, and this is true. Any reader whose interest is aroused might consult Choquet (1966).

1.4 Continuity

In this section we review the way in which a precise definition of continuity is derived from the intuitive notion. We first make a false start.

One statement containing something of the intuitive idea of continuity is that a function is continuous if its graph can be drawn without lifting pencil from paper. To formulate this more mathematically, let $f:\mathbf{R}\to\mathbf{R}$ be a function and let $(a, f(a))$, $(b, f(b))$ be two points on its graph (see Fig. 1.2). Let L be the horizontal line at some height d between $f(a)$ and $f(b)$. Then to satisfy our intuition about continuity, the graph of f has to cross the line L at least once on its way from $(a, f(a))$ to $(b, f(b))$. In other words, there exists at least one point c in $[a, b]$ such that $f(c)=d$. Formally, we make the following definition.

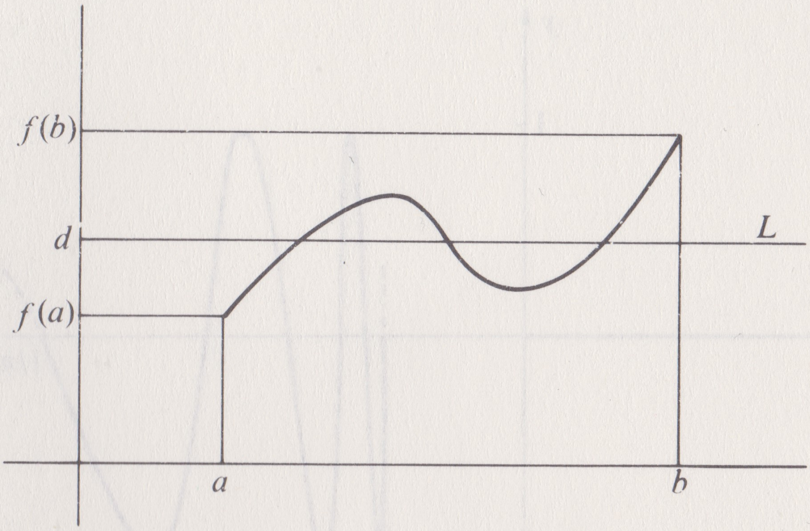


FIG. 1.2

DEFINITION 1.4.1. A function $f:\mathbf{R}\to\mathbf{R}$ has the *intermediate value property* (I.V.P.) if given any a, b, d in $\mathbf{R}$ with $a<b$ and d between $f(a)$ and $f(b)$, there exists at least one c in $\mathbf{R}$ satisfying $a\leqslant c\leqslant b$ and $f(c)=d$.

A tentative definition of continuity is that f is continuous if it has the I.V.P. However, this fails to capture completely the intuitive idea of continuity, as the next example shows.

EXAMPLE 1.4.2. Let $f:\mathbf{R}\to\mathbf{R}$ be given by

$$f(x)=\begin{cases} 0, & x\leqslant 0,\\ \sin 1/x, & x>0.\end{cases}$$

Part of the graph of f is shown in Fig. 1.3.

Although we shall not prove it now, it is easy to believe by inspection that f does have the I.V.P. But f does not satisfy our intuitive requirements for a continuous function—something is wrong near $x=0$. On closer scrutiny, we realize that our intuition includes the requirement that for all values of x near 0, $f(x)$ should be reasonably close to $f(0)$, not oscillating with amplitude 1 as it does in this example. More precisely, the reason we are dissatisfied with f is that $\lim\limits_{x\to 0} f(x)$ does not exist. Considerations such as these lead to the accepted definition.

DEFINITION 1.4.3. A function $f:\mathbf{R}\to\mathbf{R}$ is *continuous at a* if $\lim\limits_{x\to a} f(x)$ exists and equals $f(a)$. Using Definition 1.3.1 this translates into $\varepsilon-\delta$ form.

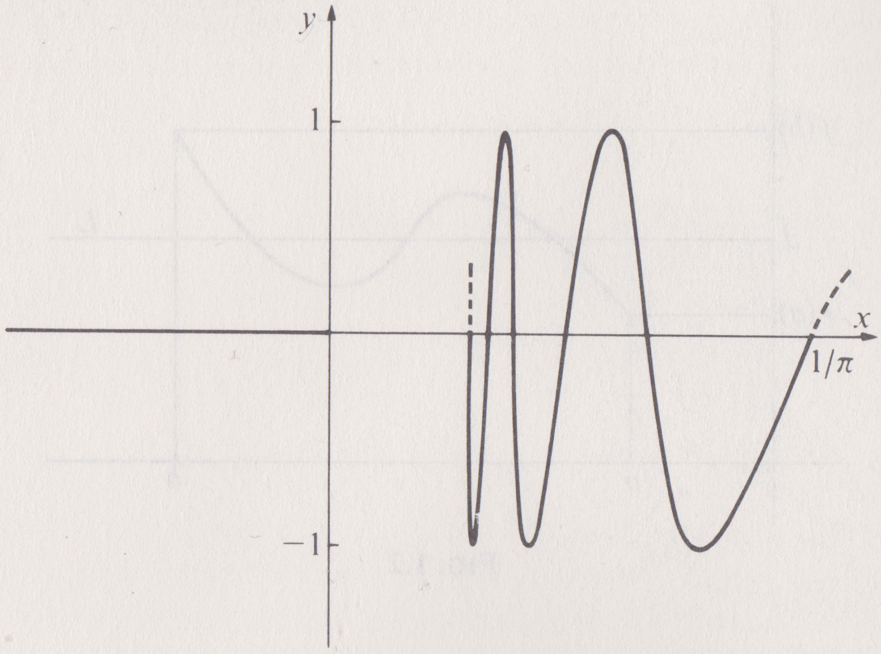


FIG. 1.3

DEFINITION 1.4.3′. A function $f: \mathbf{R} \to \mathbf{R}$ is *continuous at a* if given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ for any x satisfying $|x - a| < \delta$.

A third description of continuity of f at a, which suggests all the ways in which continuity might fail, is as follows: $\lim_{x \to a-} f(a)$ and $\lim_{x \to a+} f(x)$ both exist and equal $f(a)$.

Here are two ways of thinking about continuity of f at a.

(1) In terms of approximations: we can ensure that $f(x)$ approximates $f(a)$ within any prescribed degree of accuracy by choosing x to approximate a sufficiently accurately.

(2) Geometrically: given a horizontal band of any positive width 2ε centred on height $f(a)$, we can choose a vertical band of some suitable width 2δ centred on $x = a$ such that the part of the graph of f in this vertical band is also in the horizontal band (see Fig. 1.4; if an aeroplane is flying at 10 000 ft at time $t = a$, then it is between 11 000 ft and 9000 ft for a non-zero time interval around $t = a$, unless it is capable of discontinuous flight).

As in the definition of $\lim_{x \to a} f(x)$, it is not necessary for f to be defined on all of $\mathbf{R}$ for the definition of continuity of f at a to make sense. It is certainly enough for f to be defined on some open interval containing a, and later we shall relax this assumption further. For example we say that f is *continuous at a from the right* (*left*) if $\lim_{x\to a+} f(x)$ $\left(\lim_{x\to a-} f(x)\right)$ exists and equals $f(a)$.

DEFINITION 1.4.4. A function $f:\mathbf{R}\to\mathbf{R}$ is *continuous* if it is continuous at a for all a in $\mathbf{R}$.

It follows from the corresponding results on limits of functions that continuity behaves well under addition, multiplication, and (with suitable reservations) division of real-valued functions. Using this, it is easy to check that any polynomial function p given by $p(x)=a_0+a_1x+\ldots+a_nx^n$ is continuous, and that any rational function $p(x)/q(x)$ (where p and q are polynomials) is continuous except possibly at zeros of q—thus providing an initial fund of examples of continuous functions.

In connection with the false start we made at the beginning of this section, the following theorem is true.

THEOREM 1.4.5. *Any continuous function* $f:\mathbf{R}\to\mathbf{R}$ *has the I.V.P.*

We could give the proof now, using the completeness axiom. However, before proving this and other basic results on continuity, we shall raise the stakes by generalizing to functions between more

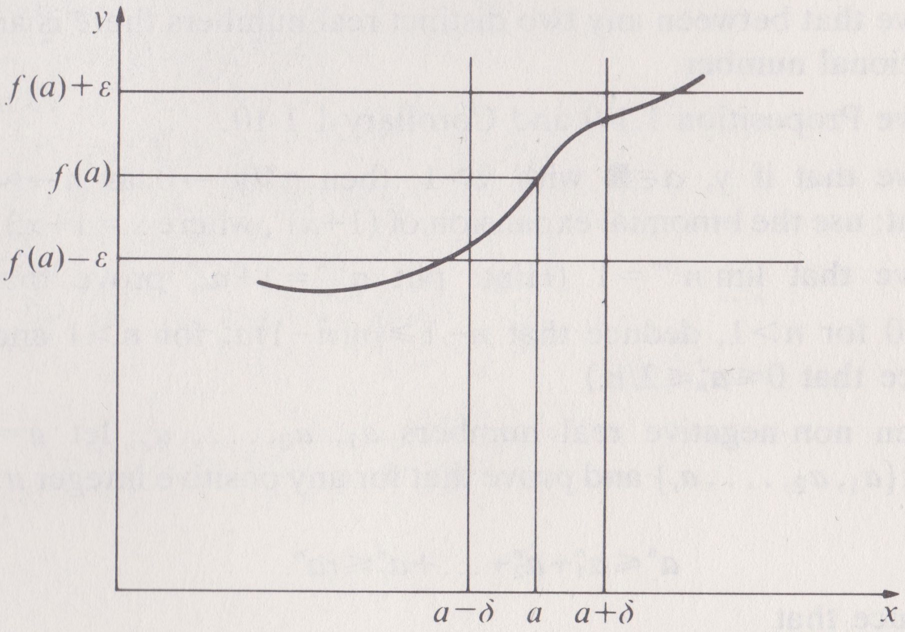


FIG. 1.4

general 'spaces' than subsets of **R**. The motives for this have been mentioned in the introduction.

Exercise 1.5

1. Show that if $A \subset B \subset \mathbf{R}$ and if B is bounded above, then A is bounded above and $\sup A \leqslant \sup B$.
2. Show that if A and B are subsets of **R** which are bounded above, then $A \cup B$ is bounded above and
$$\sup A \cup B = \operatorname{Max}\{\sup A, \sup B\}.$$
3. Formulate and prove analogues of Exercises 1 and 2 for inf.
4. For each of the following subsets of **R**, find the sup if it exists, and decide whether it is in the set:
$$\{x : x^2 \leqslant 2x - 1\},$$
$$\{x : x^2 + 2x \leqslant 1\},$$
$$\{x : x^3 < 8\},$$
$$\{x : x \sin x < 1\}.$$
5. Show that if m and n are positive integers with highest common factor 1, then m/n is the square of a rational number iff m and n are both squares of integers.
6. Prove Proposition 1.1.5 in detail.
7. Prove that between any two distinct real numbers there is an irrational number.
8. Prove Proposition 1.1.9 and Corollary 1.1.10.
9. Prove that if $y, \alpha \in \mathbf{R}$ with $y > 1$, then $n^\alpha / y^n \to 0$ as $n \to \infty$. (Hint: use the binomial expansion of $(1+x)^n$, where $y = 1 + x$.)
10. Prove that $\lim_{n \to \infty} n^{1/n} = 1$. (Hint: put $n^{1/n} = 1 + a_n$, prove that $a_n > 0$ for $n > 1$, deduce that $n - 1 \geqslant \frac{1}{2} n(n-1) a_n^2$ for $n > 1$ and hence that $0 \leqslant a_n^2 \leqslant 2/n$.)
11. Given non-negative real numbers $a_1, a_2, \ldots, a_r$, let $a = \operatorname{Max}\{a_1, a_2, \ldots, a_r\}$ and prove that for any positive integer n,
$$a^n \leqslant a_1^n + a_2^n + \ldots + a_r^n \leqslant r a^n.$$
Deduce that
$$\lim_{n \to \infty} (a_1^n + a_2^n + \ldots + a_r^n)^{1/n} = a.$$

12. Prove that if the subsequences (s_{2n}) and (s_{2n+1}) of a sequence (s_n) of real numbers both converge to the same limit l then (s_n) converges to l.

13. If (x_n) is a sequence in $\mathbf{R}$ such that $0<x_1<1$, and for all $n>1$, $x_{n+1}=2/(x_n+1)$, prove that (x_{2n}) and (x_{2n+1}) are monotonic, one decreasing and the other increasing. Deduce that (x_n) converges.

14. Let $x_1=\sqrt{2}$, $x_{n+1}=\{2+(x_n)^{\frac{1}{2}}\}^{\frac{1}{2}}$ for $n\geqslant 1$. Prove that (x_n) converges to a root of $x^4-4x^2-x+4=0$ between $\sqrt{3}$ and 2.

15. Give an example in which $f(x)\to b$ as $x\to a$ and $g(y)\to c$ as $y\to b$, but $g(f(x))\not\to c$ as $x\to a$.

16. Give the proofs necessary for Example 1.3.4.

17. Prove that if $f, g:\mathbf{R}\to\mathbf{R}$ are continuous at a, then so are $f+g$, $f-g$ and $|f|$ given by

$$(f+g)(x)=f(x)+g(x), \qquad (f-g)(x)=f(x)-g(x),$$
$$|f|\,(x)=|f(x)|.$$

18. (a) Prove that for any y, z in $\mathbf{R}$,

$$\text{Max}\{y, z\}=\tfrac{1}{2}(y+z+|y-z|),$$
$$\text{Min}\{y, z\}=\tfrac{1}{2}(y+z-|y-z|).$$

(b) Given that $f, g:\mathbf{R}\to\mathbf{R}$ are continuous at a, prove that h and k are continuous at a, where for any x in $\mathbf{R}$,

$$h(x)=\text{Max}\{f(x), g(x)\}, \qquad k(x)=\text{Min}\{f(x), g(x)\}.$$

19. Prove that if $f, g:\mathbf{R}\to\mathbf{R}$ are continuous at a then so is $f.g$ given by $(f.g)(x)=f(x)g(x)$, and that, provided $g(a)\neq 0$, the function f/g given by $(f/g)(x)=f(x)/g(x)$ is also continuous at a.

20. Let $f:\mathbf{R}\to\mathbf{R}$ be given by

$$f(x)=\begin{cases}0, & x\in\mathbf{Q},\\ 1, & x\notin\mathbf{Q}.\end{cases}$$

Show that f is not continuous at any point in $\mathbf{R}$.

21*. Let $f:\mathbf{R}\to\mathbf{R}$ be given by

$$f(x)=\begin{cases}0 \text{ if} & x=0 \quad\text{or}\quad x\notin\mathbf{Q},\\ 1/q \text{ if} & x=p/q \text{ where } p, q \text{ are integers}\\ & \text{with highest common factor 1 and } q>0.\end{cases}$$

Prove that f is discontinuous at any non-zero a in **Q**, but continuous at 0 and at any irrational a.

22*. A function $f:\mathbf{R}\to\mathbf{R}$ is said to be *convex* if

$$f(\lambda x+(1-\lambda)y)\leqslant\lambda f(x)+(1-\lambda)f(y)$$

for all x, y in **R** and all λ in $[0, 1]$. Prove that any convex function is continuous.

2

Continuity generalized: metric spaces

2.1 Motivation

WE begin by rephrasing Definition 1.4.3′ using more English and no Greek: a function $f: \mathbf{R} \to \mathbf{R}$ is continuous at a if we can make the distance between $f(x)$ and $f(a)$ (namely $|f(x)-f(a)|$) as small as we please by choosing x so that the distance between x and a (namely $|x-a|$) is sufficiently small. (The reader is reminded that the terms function and map are interchangable. We tend to use the former when dealing with functions of real variables, for which this terminology is long established, and the latter when dealing with maps between more general sets.) Next let us consider a real-valued function f of two real variables. We again get a definition corresponding to our intuitive idea of continuity by changing the above wording very slightly: f is continuous at a point (a, b) in $\mathbf{R}^2$ if we can make the distance between $f(x, y)$ and $f(a, b)$ as small as we please by choosing (x, y) so that the distance between (x, y) and (a, b) is sufficiently small. We may recover an ε–δ form by using formulae for the distances involved. Since $f(x, y)$ and $f(a, b)$ are real numbers (f is real-valued) the distance between them is $|f(x, y)-f(a, b)|$. Since (x, y) and (a, b) are points in the plane, the distance between them is $\surd\{(x-a)^2+(y-b)^2\}$, where as always this means the non-negative square root. Thus $f: \mathbf{R}^2 \to \mathbf{R}$ is continuous at (a, b) if given any $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x, y)-f(a, b)| < \varepsilon$ for all (x, y) satisfying $\surd\{(x-a)^2+(y-b)^2\} < \delta$.

As before, we may think of this definition in terms of accuracy of approximation, or geometrically. In the geometric picture, the graph of f is now some surface in three-dimensional coordinate space (at least for 'well-behaved' f), and f is continuous at (a, b) if when we choose any horizontal slab of space centred on height $f(a, b)$, there is a circular cylinder centred on the vertical line through (a, b) such that the part of the surface inside this cylinder is also trapped in the given horizontal slab. It may or may not be found helpful to draw and contemplate the appropriate picture; it is not essential for further progress.

Now, given any positive integer n, let us try to define continuity for a real-valued function f of n real variables, $f:\mathbf{R}^n\to\mathbf{R}$. We shall denote a point in $\mathbf{R}^n$ by $x=(x_1, x_2, \ldots, x_n)$. Traditionally the x_i are called the independent variables. By analogy, we may try writing: f is continuous at $a=(a_1, a_2, \ldots, a_n)$ if the distance between $f(x)$ and $f(a)$ can be made as small as we please by choosing x so that the distance between x and a is sufficiently small. Let us see if this means anything. Since $f(x)$ and $f(a)$ are real numbers, the distance between them is $|f(x)-f(a)|$. However, the definition does not make sense until we find a suitable meaning for 'the distance between points x and a in $\mathbf{R}^n$' for general n in $\mathbf{N}$.

Such a meaning is not hard to guess when we look at the particular cases $n=1, 2, 3$. When $n=3$ and we think of x and a as points in Euclidean 3-space, the distance between them is $\sqrt{\{(x_1-a_1)^2+(x_2-a_2)^2+(x_3-a_3)^2\}}$. We have already used similar formulae for $n=1, 2$. (Note that $|x-a|=\sqrt{\{(x-a)^2\}}$.) It is therefore plausible to define 'the Euclidean distance between $(x_1, x_2, \ldots, x_n)$ and $(a_1, a_2, \ldots, a_n)$' to be

$$\left\{\sum_{i=1}^{n}(x_i-a_i)^2\right\}^{\frac{1}{2}}.$$

Definition 2.1.1. A function $f:\mathbf{R}^n\to\mathbf{R}$ is *continuous at* $a=(a_1, a_2, \ldots, a_n)$ if given $\varepsilon>0$, there exists $\delta>0$ such that $|f(x)-f(a)|<\varepsilon$ for all $x=(x_1, x_2, \ldots, x_n)$ satisfying

$$\left\{\sum_{i=1}^{n}(x_i-a_i)^2\right\}^{\frac{1}{2}}<\delta.$$

For general n, this definition does not have an intuitive geometric interpretation except by analogy with the cases $n=1, 2$. However it still has familiar physical interpretations. It often happens that some physical quantity depends on several variables. For example the energy of a given solid body in a gravitational field depends on its height, its linear velocity, and its angular velocity, and these may be described by seven real variables. Continuity in the sense of Definition 2.1.1 for the function giving the energy in terms of these seven variables means that if the variables are all altered slightly the energy changes only slightly.

Now let us try generalizing one step further. The way in which we arrived at Definition 2.1.1 suggests a plausible definition of continuity for any map $f: A \to B$ provided that we can give an adequate meaning to 'the distance between' any two elements in A or in B. Roughly, a set A together with such an adequate meaning is called a metric space. Formally we consider a set A and a map $d: A \times A \to \mathbf{R}$, called a distance function or metric, so that given x, y in A, the real number $d(x, y)$ will be called the distance between x and y. Now an arbitrary map d would not give rise to an interesting theory. The properties which d has to satisfy before it is called a metric for A have evolved gradually. However, with hindsight we may abstract these properties from the particular case of the Euclidean distance in $\mathbf{R}^n$. First, $d(x, y)$ should be non-negative for any x, y in A, and it should be zero only if the 'points' x and y coincide. Formally,

$$\text{(M1)} \quad d(x, y) \geqslant 0; \qquad d(x, y) = 0 \Leftrightarrow x = y.$$

Secondly, the distance should be the same whether we 'measure' from x to y or from y to x. Formally,

$$\text{(M2)} \quad d(x, y) = d(y, x) \quad \text{for all } x, y \text{ in } A.$$

The third property of Euclidean distance which we abstract is called the triangle inequality. For the Euclidean plane or three-space it says that the sum of the lengths of two sides of any triangle is not less than the length of the third side. Formally,

$$\text{(M3)} \quad d(x, y) + d(y, z) \geqslant d(x, z) \quad \text{for all } x, y, z \text{ in } A.$$

To see that (M3) is the formal version of the triangle inequality, we think of the points x, y, z as the vertices of a 'triangle' in A.

Here then are three properties which hold for ordinary Euclidean distance in the plane and in three-space (and, rather trivially, in the real line) and do not offend geometric intuition of what distance should mean. Moreover, experience has shown that in many proofs about continuity for maps between Euclidean spaces, it is really only these three properties which are used, and not the explicit formulae for Euclidean distances. We therefore formalize this into the definition of a metric space.

DEFINITION 2.1.2. A *metric space* $M = \{A, d\}$ consists of a non-empty set A together with a map $d: A \times A \to \mathbf{R}$ satisfying (M1), (M2), (M3) above.

The elements of A are called the points of the space, and d is called a metric on A. We often loosely use the same name for M and A, when it is unmistakable which metric d is intended.

DEFINITION 2.1.3. Given two metric spaces $M_1=\{A_1, d_1\}$, $M_2=\{A_2, d_2\}$, a map $f: A_1 \to A_2$ and a point a in A_1, f is *continuous at a* (with respect to the metrics d_1, d_2) if given any $\varepsilon>0$, there exists $\delta>0$ such that $d_1(x, a)<\delta$ implies $d_2(f(x), f(a))<\varepsilon$. When necessary for clarity we say that f is (d_1, d_2)-continuous at a.

If f is continuous in this sense at every a in A_1 we say that f is *continuous.*

2.2 Examples

The extent to which the above definitions give a fruitful generalization depends largely on how many interesting examples there are of metric spaces. This section contains several examples. Some are designed to illustrate phenomena internal to metric space theory, but others are of interest in analysis.

The first example is the one from which we have abstracted the axioms.

EXAMPLE 2.2.1. *Euclidean n-space.* Let $A=\mathbf{R}^n$, and for $x=(x_1, x_2, \dots, x_n)$, $y=(y_1, y_2, \dots, y_n)$ put

$$d(x, y)=\left\{\sum_{i=1}^{n} (x_i-y_i)^2\right\}^{\frac{1}{2}}.$$

It is easy to see that (M1) and (M2) are satisfied. To check (M3), let $z=(z_1, z_2, \dots, z_n)$, let all summations be over $i=1, 2, \dots, n$, and write $x_i-y_i=r_i$, $y_i-z_i=s_i$. Then we have to prove

$$\{\sum r_i^2\}^{\frac{1}{2}}+\{\sum s_i^2\}^{\frac{1}{2}}\geqslant\{\sum (r_i+s_i)^2\}^{\frac{1}{2}}.$$

Since both sides are non-negative, it is equivalent (squaring both sides) to prove

$$\sum r_i^2+\sum s_i^2+2(\sum r_i^2)^{\frac{1}{2}}(\sum s_i^2)^{\frac{1}{2}}\geqslant\sum r_i^2+\sum s_i^2+2\sum r_i s_i.$$

This in turn is equivalent to proving Cauchy's inequality,

$$\sum r_i^2 \sum s_i^2\geqslant(\sum r_i s_i)^2,$$

one proof of which runs as follows. For any real number λ, $\sum (r_i+\lambda s_i)^2\geqslant 0$, since the left-hand side is a sum of squares of real numbers. Hence

$$f(\lambda)\equiv\sum r_i^2+2\lambda\sum r_is_i+\lambda^2\sum s_i^2\geqslant 0 \quad \text{for all real } \lambda.$$

Therefore the quadratic equation $f(\lambda)=0$ in λ cannot have two distinct real roots λ_1 and λ_2, otherwise $f(\lambda)$ would be strictly negative for λ between λ_1 and λ_2. (This follows by algebra, since we should then have

$$f(\lambda)\equiv(\lambda-\lambda_1)(\lambda-\lambda_2)\sum s_i^2.)$$

Hence the discriminant 'b^2-4ac' of $f(\lambda)$ must be non-positive. This gives Cauchy's inequality, since here $b=2\sum r_is_i$, $a=\sum s_i^2$, $c=\sum r_i^2$.

When we consider $\mathbf{R}^n$ as a metric space, the Euclidean metric will always be understood unless some other is specified.

The next example is a much stranger one.

EXAMPLE 2.2.2. *Discrete metric spaces.* Let A be any non-empty set and define d by

$$d(x,y)=\begin{cases}1, & x\neq y\\ 0, & x=y.\end{cases}$$

Then (M1) and (M2) clearly hold. To check (M3), note that if $x=z$ then $d(x,z)=0$ and $d(x,y)+d(y,z)\geqslant d(x,z)$. If $x\neq z$, then at least one of $x\neq y$, $y\neq z$ must hold, so $d(x,y)+d(y,z)$ is 1 or 2, while $d(x,z)=1$, and the triangle inequality again holds.

For any set A, this metric is called the discrete metric. (In some textbooks it is called the standard discrete metric, and the term discrete metric is applied to a wider class of metrics.) Such 'pathological' examples, as they are nicknamed, are not normally used in applications in analysis. They serve as a warning to check by rigorous proofs that results suggested by geometric intuition really hold in general metric spaces. In other words, they are potential counterexamples. Nevertheless, metric spaces built up from discrete metric spaces have been used in combinatorial problems, and if the building process is complex enough, the spaces again become of interest in analysis. The reader interested in pursuing these cryptic statements further is referred to Chapter 4 of Ash (1965) and p. 165 of Kelley (1955) respectively.

The third set of examples shows that there may be several distinct metric spaces with the same underlying set.

EXAMPLES 2.2.3. Let $A=\mathbf{R}^2$, and for $x=(x_1, x_2)$, $y=(y_1, y_2)$ let

$$d_1(x, y)=|x_1-y_1|+|x_2-y_2|,$$
$$d_2(x, y)=\{(x_1-y_1)^2+(x_2-y_2)^2\}^{\frac{1}{2}},$$
$$d_\infty(x, y)=\mathrm{Max}\{|x_1-y_1|, |x_2-y_2|\},$$

(the choice of subscripts will be explained by Example 2.2.18 and Exercise 1.5.11). We already know that d_2 satisfies the metric space axioms. It is clear that d_1 and d_∞ satisfy (M1) and (M2). To check (M3), let $z=(z_1, z_2)$. Then

$$\begin{aligned} d_1(x, y)+d_1(y, z)&=|x_1-y_1|+|x_2-y_2|+|y_1-z_1|+|y_2-z_2| \\ &\geqslant |x_1-z_1|+|x_2-z_2|=d_1(x, z). \end{aligned}$$

Also, for $i=1, 2$, we have

$$|x_i-z_i|\leqslant|x_i-y_i|+|y_i-z_i|\leqslant d_\infty(x, y)+d_\infty(y, z),$$

so

$$d_\infty(x, z)\leqslant d_\infty(x, y)+d_\infty(y, z).$$

We could also let d_0 be the discrete metric on $\mathbf{R}^2$. We shall see in § 2.4 that d_1, d_2, d_∞ are all equivalent in a certain sense, but that they are not equivalent to d_0.

It is clear how to define analogues of these metrics on $\mathbf{R}^n$ for any n in $\mathbf{N}$. The proofs that the axioms hold are similar to the above.

EXAMPLE 2.2.4. Let $A=\mathbf{C}$, and for z_1, z_2 in $\mathbf{C}$ let $d(z_1, z_2)=|z_1-z_2|$. Again (M1) and (M2) clearly hold. When we express each complex number in terms of its real and imaginary parts, the triangle inequality $|z_1-z_2|+|z_2-z_3|\geqslant|z_1-z_3|$ for $\mathbf{C}$ coincides with the triangle inequality for $\mathbf{R}^2$. The close relation between this metric space and $\mathbf{R}^2$ will be made precise in § 2.4.

Examples 2.2.5 and 2.2.7, subspaces and product spaces, have counterparts for many other mathematical structures, for example, in algebra.

EXAMPLE 2.2.5. *Metric subspaces.* Suppose that $M=\{A, d\}$ is a metric space and that H is a non-empty subset of A. Let $d_H: H\times H\to\mathbf{R}$ be the restriction $d\,|\,H\times H$ (recall that this means $d_H(x, y)=d(x, y)$ for any x, y in H). The metric space axioms hold for d_H since they hold for d. The metric space $\{H, d_H\}$ is called a subspace of M and d_H is called the metric on H induced by d. Since the same name is loosely used for a metric space and its underlying set, H will be called a subset of M or a subspace of M interchangably, according to the emphasis desired at the time. If $H \subset \mathbf{R}^n$, then in referring to H as a metric space we shall always assume that the metric is induced by the Euclidean metric on $\mathbf{R}^n$, unless some other is specified. In particular this applies to the various kinds of intervals in $\mathbf{R}$.

When we have metric spaces $\{A, d\}$, $\{A', d'\}$, and a map $f: A \to A'$, then for $h \in H \subset A$ we can talk about continuity (more precisely (d_H, d')-continuity) of $f|H$ at h. It is important to distinguish between this and continuity of f at h.

EXAMPLES 2.2.6. As an extreme example, consider $f: \mathbf{R} \to \mathbf{R}$ given by

$$f(x) = \begin{cases} 0, & x \in \mathbf{Q}, \\ 1, & x \notin \mathbf{Q}. \end{cases}$$

Then $f|\mathbf{Q}: \mathbf{Q} \to \mathbf{R}$ is the constant function with value 0, and is continuous at every point in $\mathbf{Q}$, whereas f is not continuous at any point. For another example which will be familiar to readers who have studied real-valued functions of two real variables, let $f: \mathbf{R}^2 \to \mathbf{R}$ be given by

$$f(0, 0) = 0, \qquad f(x_1, x_2) = \frac{x_1 x_2}{x_1^2 + x_2^2} \quad \text{otherwise.}$$

Then $f|\mathbf{R} \times \{0\}$ and $f|\{0\} \times \mathbf{R}$ are both constant functions with value 0 and hence are continuous, but f is not continuous at $0 = (0, 0)$. To see this, let $\varepsilon = \frac{1}{2}$. For any $\delta > 0$, the point $x = (\frac{1}{2}\delta, \frac{1}{2}\delta)$ satisfies $d_2(x, 0) < \delta$ but $|f(x) - f(0)| = \frac{1}{2}$. This is traditionally described by saying that f is continuous in each variable separately, but not in both variables jointly, at 0.

These examples might suggest that continuity of $f|H$ is not very useful. However, suppose that a real-valued function f is defined on some subset of $\mathbf{R}$ containing $[a, b]$. Let us see what continuity of $f|[a, b]$ means. At the point a it means: given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ for all x satisfying $|x - a| < \delta$ *and* x *in* $[a, b]$, or equivalently for all x satisfying $a \leqslant x < a + \delta$. This means continuity of f from the right at a. Similarly at b it means continuity from the left. Finally, for any c in (a, b) it means ordinary (two-sided) continuity. Thus continuity of $f|[a, b]$ *is* of interest, as the reader who has studied the mean value theorem in differential calculus knows.

EXAMPLE 2.2.7. *Product spaces.* Given two metric spaces $M=\{A, d\}$, $M'=\{A', d'\}$, we may define several metrics on $A\times A'$, as in Example 2.2.3. If $x=(x_1, x_2)$ and $y=(y_1, y_2)$ are in $A\times A'$, let

$$d_1(x, y)=d(x_1, y_1)+d'(x_2, y_2),$$
$$d_2(x, y)=[\{d(x_1, y_1)\}^2+\{d'(x_2, y_2)\}^2]^{\frac{1}{2}},$$
$$d_\infty(x, y)=\mathrm{Max}\{d(x_1, y_1), d'(x_2, y_2)\}.$$

These may be proved to be metrics just as in the case $A=A'=\mathbf{R}$. Any one of them deserves to be called a product metric. In § 2.4 we shall see that they are all equivalent in a certain sense. This definition may be extended to the product of any finite number of metric spaces.

If Examples 2.2.1–2.2.4, together with subspaces and products formed from them, were the only examples, it is doubtful whether general metric space theory would be very profitable. The examples below may help to remove that doubt. They are examples of function spaces: we take some collection of functions and decide to treat it as a 'space', calling the individual functions 'points' and (in our case) putting a metric on the collection. To give some indication of how this might be useful, let us consider a classical problem in one of the first areas where function space language was seen to be appropriate, the calculus of variations. The brachistochrone problem is roughly as follows. Suppose we have any two points x, y in a vertical plane, with x higher than, but not vertically above, y. What is the shape of the curve in this plane along which a heavy particle will take the least time to slide from x to y under the action of gravity (with no friction)? It is not essential for the purposes of this illustration to be precise about what kind of curves are intended, but for example we could take 'curve' to mean one given by a function $h:[0, 1]\to\mathbf{R}^2$, say $h(t)=(f(t), g(t))$, where $f, g:[0, 1]\to\mathbf{R}$ are continuously differentiable functions. For any given curve λ we can use integration to calculate the time $T(\lambda)$ the particle takes to make the journey from x to y along λ. Thus we get a real-valued function T defined on the collection of all curves from x to y, and the brachistochrone problem is to find the 'point' λ_0 (if there is one) at which T takes a minimum value. (The answer turns out to be part of a cycloid.) This is like looking for a minimum of a function $T:\mathbf{R}\to\mathbf{R}$, except that the domain $\mathbf{R}$ is replaced by the 'space' of curves. The calculus of variations develops analogues of ordinary calculus for solving such problems. Even to begin calculus, we need continuity of T, and this motivates putting a metric on the collection of curves. The latter collection is more or less the set of functions h of the type mentioned above satisfying $h(0)=x$, $h(1)=y$, except for a technicality concerning the fact that distinct functions may describe the same geometric curve (for example, $h_1, h_2:[0,1]\to\mathbf{R}^2$ given by $h_1(t)=(t,t)$ and $h_2(t)=(t^2,t^2)$ both describe a straight line segment joining the points $(0,0)$ and $(1,1)$).

In this and other contexts where maps are defined on collections of functions, the language of function spaces is useful. The possibilities it allows for the use of geometric intuition have proved fruitful.

EXAMPLE 2.2.8. Let A be the set of all bounded functions $f:[a,b]\to\mathbf{R}$. Given two points f and g in A, let

$$d(f,g)=\sup_{x\in[a,b]}|f(x)-g(x)|.$$

The right-hand side exists, since if f and g are bounded, say $|f(x)|\leqslant K$, $|g(x)|\leqslant L$ for all x in $[a,b]$, then

$$|f(x)-g(x)|\leqslant|f(x)|+|g(x)|\leqslant K+L \quad \text{for all } x \text{ in } [a,b].$$

We shall check in detail that the metric space axioms hold.

(M1) It is clear that $d(f,g)\geqslant 0$, since it is the sup of a set of non-negative numbers. Also, if f and g are the same point of A, this means that they are identical as functions from $[a,b]$ to $\mathbf{R}$, so $f(x)=g(x)$ for all x in $[a,b]$, and $\sup_{x\in[a,b]}|f(x)-g(x)|=0$. Conversely this sup is zero only if $|f(x)-g(x)|=0$ for every x in $[a,b]$, which implies that f and g are identical as functions; in other words they are equal as points of A.

(M2) Since $|f(x)-g(x)|=|g(x)-f(x)|$ for all x in $[a,b]$, it follows that $\sup_{x\in[a,b]}|f(x)-g(x)|=\sup_{x\in[a,b]}|g(x)-f(x)|$, as required.

(M3) Let $f, g, h\in A$. For any c in $[a,b]$,

$$\begin{aligned}|f(c)-h(c)|&\leqslant|f(c)-g(c)|+|g(c)-h(c)|\\&\leqslant\sup_{x\in[a,b]}|f(x)-g(x)|+\sup_{x\in[a,b]}|g(x)-h(x)|\\&=d(f,g)+d(g,h),\end{aligned}$$

where the first inequality is just the triangle inequality for $\mathbf{R}$. The above holds for any c in $[a, b]$, so $d(f, g)+d(g, h)$ is an upper bound for the set $S=\{|f(c)-h(c)| : c\in[a, b]\}$. Hence $d(f, g)+d(g, h)\geqslant \sup S = d(f, h)$, as required.

This metric is called the *sup* metric or the *uniform* metric. We denote the resulting metric space by $\mathscr{B}([a, b], \mathbf{R})$; the notation is not universally agreed.

By the theorem quoted in the introduction, any continuous $f:[a, b]\to\mathbf{R}$ is bounded. The set of all such continuous functions therefore forms a subspace of $\mathscr{B}([a, b], \mathbf{R})$, sometimes denoted by $\mathscr{C}[a, b]$.

EXAMPLE 2.2.9. Let A be the set of all continuous $f:[a, b]\to\mathbf{R}$, but this time let

$$d(f, g)=\int_a^b |f(x)-g(x)|\,dx.$$

To check the metric space axioms, we need the following lemma from integration theory.

LEMMA 2.2.10. *Suppose that* $h:[a, b]\to\mathbf{R}$ *is continuous, that* $h(x)\geqslant 0$ *for all* x *in* $[a, b]$, *and that* $\int_a^b h(x)\,dx=0$. *Then* $h(x)=0$ *for all* x *in* $[a, b]$.

The idea of the proof is that if $h(c)>0$ for some c in $[a, b]$, then by continuity $h(x)$ exceeds some fixed positive number (for example, $\frac{1}{2}h(c)$) throughout an interval of non-zero length. This makes a positive contribution to the integral which cannot be cancelled out elsewhere since $h(x)$ is never negative.

(M1) It is clear that $d(f, g)\geqslant 0$ and that if $f=g$ then $d(f, g)=0$. If $d(f, g)=0$ then, by Lemma 2.2.10 applied with $h=|f-g|$, $f=g$.

(M2) Symmetry of $d(f, g)$ is clear.

(M3) For any continuous f, g, $h:[a, b]\to\mathbf{R}$ and any x in $[a, b]$,

$$|f(x)-g(x)|+|g(x)-h(x)|\geqslant|f(x)-h(x)|.$$

Hence, by integration theory,

$$\int_a^b |f(x)-g(x)|\,dx+\int_a^b |g(x)-h(x)|\,dx\geqslant\int_a^b |f(x)-h(x)|\,dx,$$

as required. This metric is called the L^1 metric.

Examples 2.2.8 and 2.2.9 give us a choice of two metrics on the set of continuous $f:[a, b]\rightarrow\mathbf{R}$. Let us denote them by d_∞ and d_1 respectively. The metric used in any particular situation depends on which properties of the functions are of interest at the time. If we regard g as a good approximation to f iff $g(x)$ is uniformly close to $f(x)$ for all x, we use d_∞ (this will be studied in Chapter 8). On the other hand we might not be as much interested in the difference in values of functions at each particular point as in their average deviation from one another over some range. We might then use d_1 or some other metric involving integration, such as the next example.

EXAMPLE 2.2.11. Let A be as in Example 2.2.9 and let

$$d_2(f, g)=\left\{\int_a^b (f(x)-g(x))^2\, dx\right\}^{\frac{1}{2}}.$$

Again (M1) follows from Lemma 2.2.10, and (M2) clearly holds. The proof that (M3) holds is similar to the proof in Example 2.2.1, with Cauchy's inequality replaced by its analogue for integrals, the Cauchy–Schwarz–Bunyakovsky inequality:

$$\int_a^b (f(x))^2\, dx . \int_a^b (g(x))^2\, dx \geqslant \left\{\int_a^b f(x)g(x)\, dx\right\}^2.$$

This may be proved as we proved Cauchy's inequality, beginning with the inequality

$$\int_a^b (f(x)+\lambda g(x))^2\, dx \geqslant 0$$

for all λ in $\mathbf{R}$. The metric d_2 is called the L^2 metric.

At this stage it is convenient to generalize the notion of a bounded set in $\mathbf{R}$.

DEFINITION 2.2.12. A subset S of a metric space $M=\{A, d\}$ is *bounded* if there exist a in A and K in $\mathbf{R}$ such that $d(x, a)\leqslant K$ for all x in S.

Thus a plays the role of 0 in $\mathbf{R}$. However, if S satisfies the definition for some a in A and K in $\mathbf{R}$, then it also satisfies the definition with a replaced by any other point a' in A and K

replaced by $K' = K + d(a, a')$. For if $d(x, a) \leqslant K$ then $d(x, a') \leqslant d(x, a) + d(a, a') \leqslant K + d(a, a')$.

If S satisfies Definition 2.2.12, then for any x, y in S, we have $d(x, y) \leqslant d(x, a) + d(a, y) \leqslant 2K$. The following definition therefore makes sense.

DEFINITION 2.2.13. If S is a non-empty bounded subset of a metric space with metric d, then the *diameter* diam S is sup D, where $D \subset \mathbf{R}$ is the set $\{d(x, y) : x, y \in S\}$.

DEFINITION 2.2.14. If $f : X \to M$ is a map into a metric space M, then f is *bounded* if the subset $f(X)$ of M is bounded.

PROPOSITION 2.2.15. *The union of any finite number of bounded subsets of a metric space is bounded.*

Proof. It is enough to prove this for two bounded subsets, since the proposition then follows by induction. The proof is suggested by geometric intuition. The reader is strongly encouraged to draw and contemplate two bounded subsets in the plane before proceeding. Suppose that S_1, S_2 are bounded subsets in a metric space $M = \{A, d\}$, and let a_1, a_2 in A and K_1, K_2 in $\mathbf{R}$ be such that $d(x, a_i) \leqslant K_i$ for all x in S_i $(i = 1, 2)$. Put $a = a_1$, say, and $K = \text{Max}\{K_1, K_2 + d(a_1, a_2)\}$. Then for any x in $S_1 \cup S_2$, either $x \in S_1$ and $d(x, a) \leqslant K_1 \leqslant K$, or $x \in S_2$ and $d(x, a) = d(x, a_1) \leqslant d(x, a_2) + d(a_2, a_1) \leqslant K_2 + d(a_2, a_1) \leqslant K$.

*EXAMPLE 2.2.16. *Hilbert sequence space* ℓ^2. Let A be the set of all sequences (x_r) of real numbers such that the series $\sum_1^\infty x_r^2$ converges. For $x = (x_r)$, $y = (y_r)$ in A let

$$d_2(x, y) = \left\{\sum_{r=1}^{\infty} (x_r - y_r)^2\right\}^{\frac{1}{2}}.$$

We first have to prove that $\sum_{r=1}^{\infty} (x_r - y_r)^2$ converges. More generally we shall prove that $\sum_{r=1}^{\infty} (\lambda x_r + \mu y_r)^2$ converges for any λ, μ in $\mathbf{R}$. For any n,

$$\sum_{r=1}^{n} (\lambda x_r + \mu y_r)^2 = \lambda^2 \sum_{r=1}^{n} x_r^2 + \mu^2 \sum_{r=1}^{n} y_r^2 + 2\lambda\mu \sum_{r=1}^{n} x_r y_r.$$

By Cauchy's inequality,

$$\sum_{r=1}^{n} x_r y_r \leqslant \left(\sum_{r=1}^{n} x_r^2\right)^{\frac{1}{2}} \left(\sum_{r=1}^{n} y_r^2\right)^{\frac{1}{2}}.$$

Hence

$$\begin{aligned}\sum_{r=1}^{n}(\lambda x_r+\mu y_r)^2 &\leqslant \lambda^2\sum_{r=1}^{n}x_r^2+\mu^2\sum_{r=1}^{n}y_r^2+\\ &\qquad +2\,|\lambda\mu|\left(\sum_{r=1}^{n}x_r^2\right)^{\frac{1}{2}}\left(\sum_{r=1}^{n}y_r^2\right)^{\frac{1}{2}}\\ &=\left\{|\lambda|\left(\sum_{r=1}^{n}x_r^2\right)^{\frac{1}{2}}+|\mu|\left(\sum_{r=1}^{n}y_r^2\right)^{\frac{1}{2}}\right\}^2\\ &\leqslant\left\{|\lambda|\left(\sum_{r=1}^{\infty}x_r^2\right)^{\frac{1}{2}}+|\mu|\left(\sum_{r=1}^{\infty}y_r^2\right)^{\frac{1}{2}}\right\}^2.\end{aligned}$$

This shows that the sequence of partial sums $\sum_{r=1}^{n}(\lambda x_r+\mu y_r)^2$ is bounded above, and since it is also monotonic increasing, convergence of $\sum_{r=1}^{\infty}(\lambda x_r+\mu y_r)^2$ follows by Theorem 1.2.6.

It is clear that (M1) and (M2) hold in this case. To check (M3) let $z=(z_r)$ be a third point in A. Then for any n,

$$\begin{aligned}\left\{\sum_{r=1}^{n}(x_r-z_r)^2\right\}^{\frac{1}{2}} &\leqslant\left\{\sum_{r=1}^{n}(x_r-y_r)^2\right\}^{\frac{1}{2}}+\left\{\sum_{r=1}^{n}(y_r-z_r)^2\right\}^{\frac{1}{2}}\\ &\leqslant d_2(x,y)+d_2(y,z),\end{aligned}$$

where the first inequality comes from Example 2.2.1. Since this holds for all n, (M3) follows.

EXAMPLE 2.2.17. Example 2.2.8 may be generalized in several ways. The reader may have noticed that in checking the axioms for it we nowhere used the fact that the domain of all the functions concerned was $[a, b]$. We may equally well consider the set of all bounded real-valued functions $\mathscr{B}(X, R)$ on any fixed set X, with the sup metric. For example, if $X=\mathbf{N}$, then we get the space of all bounded sequences of real numbers with the metric

$$d((x_n),(y_n))=\sup_{n\in\mathbf{N}}|x_n-y_n|.$$

Also, the space **R** in which the functions take their values may be replaced by any metric space M. The set $\mathscr{B}(X, M)$ of all bounded maps from X to M has a sup metric just as in Example 2.2.8.

So far X need not be a metric space, but if it is then we get an interesting subspace $\mathscr{C}(X, M)$ of $\mathscr{B}(X, M)$ by considering the set of all bounded continuous maps from X to M (again, the notation is variable).

In a slightly different direction, we get an interesting subset $\mathscr{D}^r[a, b]$ of $\mathscr{C}[a, b]$ by considering, for some fixed $r \geqslant 1$, those functions which are r times continuously differentiable on $[a, b]$. For any $0 \leqslant i \leqslant r$, the C^i metric on $\mathscr{D}^r[a, b]$ is defined by

$$d(f, g) = \sup_{x \in [a,b]} \{|f(x) - g(x)|, |f'(x) - g'(x)|, \ldots, |f^{(i)}(x) - g^{(i)}(x)|\},$$

where $f^{(i)}$ denotes the ith derivative of f. The proof that this is a metric is similar to that in Example 2.2.8. The metric induced on $\mathscr{D}^r[a, b]$ by the (sup) metric on $\mathscr{C}[a, b]$ is of course the C^0 or sup metric, and with the C^i metric for $i > 0$, $\mathscr{D}^r[a, b]$ is a subset but not a subspace of $\mathscr{C}[a, b]$. If f, g in $\mathscr{D}^1[0, 1]$ are given by $f(x) = 0$ and $g(x) = 10^{-6} \sin 10^{16} x$ for all x in $[0, 1]$, then the distance between f and g is 10^{-6} in the C^0 metric but 10^{10} in the C^1 metric. As usual, the metric used in any particular case depends on what is currently of interest. If the unit of length in either coordinate direction represents a kilometre, then one would not worry about riding a bicycle along a road offering the graph of g as its surface, but one should worry about paying for it if there is a charge of 1 p per 1000 km *measured along the graph* of g.*

EXAMPLE 2.2.18. The metrics d_1, d_2 in Example 2.2.3 may be generalized as follows. Let $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ be points in $\mathbf{R}^n$, and for any integer $p \geqslant 1$ let

$$d_p(x, y) = \left\{\sum_{i=1}^{n} |x_i - y_i|^p\right\}^{1/p},$$

taking the non-negative real pth root. Again (M1) and (M2) are clearly satisfied. The triangle inequality for d_p follows from a generalization of Cauchy's inequality called Minkowski's inequality. We refer to Copson (1968) or Pitts (1972) for the details. Likewise there is an L^p-metric analogous to the metrics in Examples 2.2.9 and 2.2.11 and a space ℓ^p analogous to ℓ^2 for any integer $p \geqslant 1$.

REMARK. For applications in analysis, metric spaces have a strong rival in normed vector spaces. The latter are particularly convenient for generalizing differential calculus and for handling linear problems in general (see Kolmogorov and Fomin (1957)).

2.3 Open sets in metric spaces

We now develop some terminology which is useful for discussing continuity in metric spaces.

DEFINITION 2.3.1. Given a metric space $M=\{A, d\}$, a point a in A and a strictly positive real number ε, the *open ε-ball neighbourhood of a in M* is the set

$$B_\varepsilon(a)=\{x\in A : d(x, a)<\varepsilon\}.$$

The name and notation for this vary widely; *spherical neighbourhood* and *open sphere* are also common. We shall sometimes abbreviate it to *open ball.* Our notation does not exhibit the metric; when necessary we expand it to $B_\varepsilon(a; d)$.

EXAMPLE 2.3.2. In $\mathbf{R}$, $B_\varepsilon(a)=(a-\varepsilon, a+\varepsilon)$. In $\mathbf{R}^2$, $B_\varepsilon(a)$ is the interior of a disc of radius ε centred on a. In $\mathbf{R}^3$, $B_\varepsilon(a)$ is the interior of a solid ball centred on a (in all these examples, the Euclidean metric is understood). The terminology 'open ball' arises from these cases. However, the next two examples warn us not to take the name too seriously.

EXAMPLE 2.3.3. In $\mathbf{R}^2$ with the metric d_∞ of Example 2.2.3, $B_\varepsilon(a)$ is the interior of a square centred on a, with sides of length 2ε parallel to the coordinate axes.

EXAMPLE 2.3.4. In a discrete metric space $M=\{A, d\}$,

$$B_\varepsilon(a)=\begin{cases}\{a\} & \text{if} \quad \varepsilon\leqslant 1,\\ A & \text{if} \quad \varepsilon>1.\end{cases}$$

Examples 2.3.3 and 2.3.4 show that $B_\varepsilon(a; d)$ depends in general on d. It also depends on the underlying set. This is a special case of its dependence on d, since a subspace metric is logically distinct from d.

EXAMPLE 2.3.5. Let $H=[0, 1]\subset\mathbf{R}$, with the Euclidean metric d on $\mathbf{R}$. Then $B_1(1; d)=(0, 2)$, while $B_1(1; d_H)=(0, 1]$.

The definition of continuity can be rephrased in terms of open balls.

DEFINITION 2.3.6. A map $f: M_1\rightarrow M_2$ between metric spaces is continuous at x in M_1 if given any $B_\varepsilon(f(x))$ there exists a $B_\delta(x)$ such that $f(B_\delta(x))\subset B_\varepsilon(f(x))$.

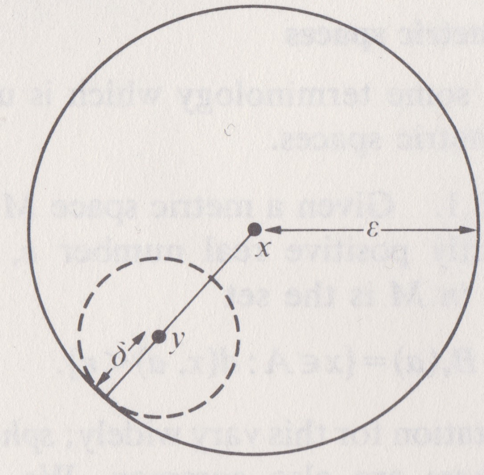


FIG. 2.1

This is just a translation of Definition 2.1.3, but it is more suitable for later use.

LEMMA 2.3.7. *Given an open ball $B_\varepsilon(x)$ in a metric space and a point y in $B_\varepsilon(x)$, there exists $\delta > 0$ such that $B_\delta(y) \subset B_\varepsilon(x)$.*

Proof. The proof illustrates how geometric intuition and analytic rigour both play a part in metric space theory. In $\mathbf{R}^2$, the conclusion asserts that we can draw a disc centred on y and lying entirely within the larger disc (see Fig. 2.1). This is obvious in the picture, but the proof for general metric spaces will have to use the metric space axioms only. However, the picture helps by suggesting what size to try taking δ, namely, such that $\delta + d(x, y) \leqslant \varepsilon$.

Here is the formal proof. Let $\delta = \varepsilon - d(x, y)$. This is strictly positive, since $y \in B_\varepsilon(x)$. We shall prove that $B_\delta(y) \subset B_\varepsilon(x)$. If $z \in B_\delta(y)$ then $d(y, z) < \delta$, and

$$d(x, z) \leqslant d(x, y) + d(y, z) < d(x, y) + \delta = \varepsilon,$$

so $z \in B_\varepsilon(x)$ as required.

The next definition generalizes that of open balls.

DEFINITION 2.3.8. A subset U of a metric space M is *open in* M if given any y in U there exists $\varepsilon(y) > 0$ such that $B_{\varepsilon(y)}(y) \subset U$.

It is important that for different points y in U the values of $\varepsilon(y)$ involved are allowed to be different.

EXAMPLES 2.3.9. By Lemma 2.3.7, any open ball in a metric space M is open in M. In particular, any 'open interval' in $\mathbf{R}$ is open in $\mathbf{R}$. On the other hand, the intervals $[a, b]$, $[a, b)$, $(a, b]$ are not open in **R**: for example $a \in [a, b)$, but no matter how small $\varepsilon > 0$ is, $B_\varepsilon(a)$ contains points of **R** (to the left of a) which are not in $[a, b)$. Not every open set is an open ball; for example, in $\mathbf{R}^2$ let U be the interior of a rectangle,

$$U = \{(x_1, x_2) : a < x_1 < b, c < x_2 < d\}.$$

Then U is not an open ball in the Euclidean metric, but U is open in $\mathbf{R}^2$, since if $x = (x_1, x_2)$ is in U, it is easily checked that $B_\varepsilon(x) \subset U$, where $\varepsilon = \text{Min}\{x_1 - a, b - x_1, x_2 - c, d - x_2\}$. As these Euclidean examples suggest, there are no 'boundary points' in a set U which is open in M, i.e. from any point in U one can 'go' some positive distance in M without going outside of U. Another way of saying the same thing is that there are no points in U from which one can get out of U by moving an arbitrarily short distance in M.

EXAMPLE 2.3.10. For any metric space M, the empty set $\varnothing$ and the whole space M are open in M. This follows trivially from the definition. For example $[a, b]$ is open in $[a, b]$. (We cannot get out of $[a, b]$ by moving an arbitrarily short distance—or indeed any distance—away from a *in* $[a, b]$, although we could in **R**.)

EXAMPLE 2.3.11. In a discrete metric space M any subset of M is open in M. For given x in U, we have only to choose $\varepsilon \leqslant 1$ to ensure that $B_\varepsilon(x) = \{x\} \subset U$.

The next group of examples show that when we say that a set is open we have to be careful what metric and what underlying set are involved.

EXAMPLE 2.3.12. A singleton set such as $\{0\}$ is open in **R** with the discrete metric but not in **R** with its usual metric. When necessary we say that a set is d-open. The interval $[a, b]$ is open in $[a, b]$ as above but not in the larger space **R**. A more illuminating example is the subset $(a, b) \times \{0\}$ of the subspace $\mathbf{R} \times \{0\}$ of $\mathbf{R}^2$. This subset is open in $\mathbf{R} \times \{0\}$, but not in the larger space $\mathbf{R}^2$. For given x in $(a, b) \times \{0\}$, any $B_\varepsilon(x)$ in $\mathbf{R}^2$ contains points which are off the x_1-axis and hence not in $(a, b) \times \{0\}$ (the reader is invited to draw a picture).

Next we derive yet another criterion for a map between metric spaces to be continuous, this time in terms of open sets. The reader may think that we are not making much progress, but merely juggling with definitions. This is true, but evidence will be offered in the next section to support the view that this criterion has some advantages over Definition 2.3.6.

PROPOSITION 2.3.13. *Suppose that $f: M_1 \to M_2$ is a map between metric spaces. Then f is continuous iff for every set U open in M_2, $f^{-1}(U)$ is open in M_1.*

Proof. Suppose that f is continuous, and let U be open in M_2. We have to show that $f^{-1}(U)$ is open in M_1. Let $x \in f^{-1}(U)$. Then $f(x) \in U$, and since U is open in M_2 there exists $\varepsilon > 0$ such that $B_\varepsilon(f(x)) \subset U$. By Definition 2.3.6 (continuity), there exists $\delta > 0$ such that $f(B_\delta(x)) \subset B_\varepsilon(f(x))$. Hence $f(B_\delta(x)) \subset U$, so $B_\delta(x) \subset f^{-1}(U)$, and we have checked that $f^{-1}(U)$ is open in M_1.

Conversely suppose that the condition on open sets holds. We shall prove that f is continuous at an arbitrary point x in M_1. For given $\varepsilon > 0$, $B_\varepsilon(f(x))$ is open in M_2 by Lemma 2.3.7, so by hypothesis $f^{-1}(B_\varepsilon(f(x)))$ is open in M_1. Also, $x \in f^{-1}(B_\varepsilon(f(x)))$ since $f(x) \in B_\varepsilon(f(x))$, so by definition of an open set, there exists $\delta > 0$ such that $B_\delta(x) \subset f^{-1}(B_\varepsilon(f(x)))$. Then $f(B_\delta(x)) \subset B_\varepsilon(f(x))$. This proves that f is continuous at x according to Definition 2.3.6.

WARNING 2.3.14. When $f: M_1 \to M_2$ is continuous, it is not necessarily true that if U is open in M_1 then the (direct) image $f(U)$ is open in M_2. For example if $f: \mathbf{R}^2 \to \mathbf{R}$ is the constant function with $f(x) = 0$ for all x, then f is continuous but for any non-empty open set U in $\mathbf{R}^2$, $f(U) = \{0\}$ which is not open in $\mathbf{R}$.

Since the process of taking inverse images of subsets under maps may not be very familiar, we give two examples.

EXAMPLE 2.3.15. Let $f: \mathbf{R} \to \mathbf{R}$ be given by

$$f(0) = 1, \qquad f(x) = x, \qquad x \neq 0.$$

Then $f^{-1}(-\infty, 0) = \{x \in \mathbf{R} : f(x) < 0\} = (-\infty, 0)$, since $f(x) < 0$ iff $x < 0$,

$$f^{-1}(0, \infty) = \{x \in \mathbf{R} : f(x) > 0\} = [0, \infty),$$

since $f(x) > 0$ iff $x > 0$ or $x = 0$. Similarly $f^{-1}(0, 1) = (0, 1)$, $f^{-1}(-1, 1) = (-1, 0) \cup (0, 1)$,

$$f^{-1}\{(-2, -1) \cup (\tfrac{1}{2}, 2)\} = (-2, -1) \cup (\tfrac{1}{2}, 2) \cup \{0\}.$$

From the viewpoint of Proposition 2.3.13, failure of continuity shows up in $f^{-1}(U)$ for any open set U containing 1 but such that there is no $\varepsilon > 0$ with $(-\varepsilon, 0) \subset U$. For then $0 \in f^{-1}(U)$, but there is no $\varepsilon>0$ such that $(-\varepsilon, 0)\subset f^{-1}(U)$, hence no $\varepsilon>0$ such that $B_\varepsilon(0)\subset f^{-1}(U)$, and $f^{-1}(U)$ is not open.

EXAMPLE 2.3.16. (a) Let $f:\mathbf{R}^2\to\mathbf{R}$ be given by $f(x, y)=xy$. Let $S=f^{-1}(1, \infty)=\{(x, y)\in\mathbf{R}^2: xy>1\}$. Then for any $x>0$, S contains those (x, y) for which $y>1/x$. Similarly for $x<0$, S contains those (x, y) with $y<1/x$. Finally, no (x, y) with $x=0$ satisfies $xy>1$. Thus S is the union of the unshaded regions in the first and third quadrants in Fig. 2.2, bounded by the hyperbola $xy=1$, and not including the boundary curves. Similarly, $f^{-1}(0, 1)$ is the shaded region in Fig. 2.2, not including the boundary curves and lines.

(b) Let $g:\mathbf{R}^2\to\mathbf{R}^2$ be given by $g(x, y)=(x^2+y^2, xy)$. Then for example

$$\begin{aligned} S &= g^{-1}\{(0, 2)\times(0, 1)\} \\ &= \{(x, y)\in\mathbf{R}^2: 0<x^2+y^2<2 \text{ and } 0<xy<1\}. \end{aligned}$$

We know from (a) that $\{(x, y)\in\mathbf{R}^2: 0<xy<1\}$ is the shaded region in Fig. 2.2. Also, $\{(x, y)\in\mathbf{R}^2: 0<x^2+y^2<2\}$ is the interior of a disc

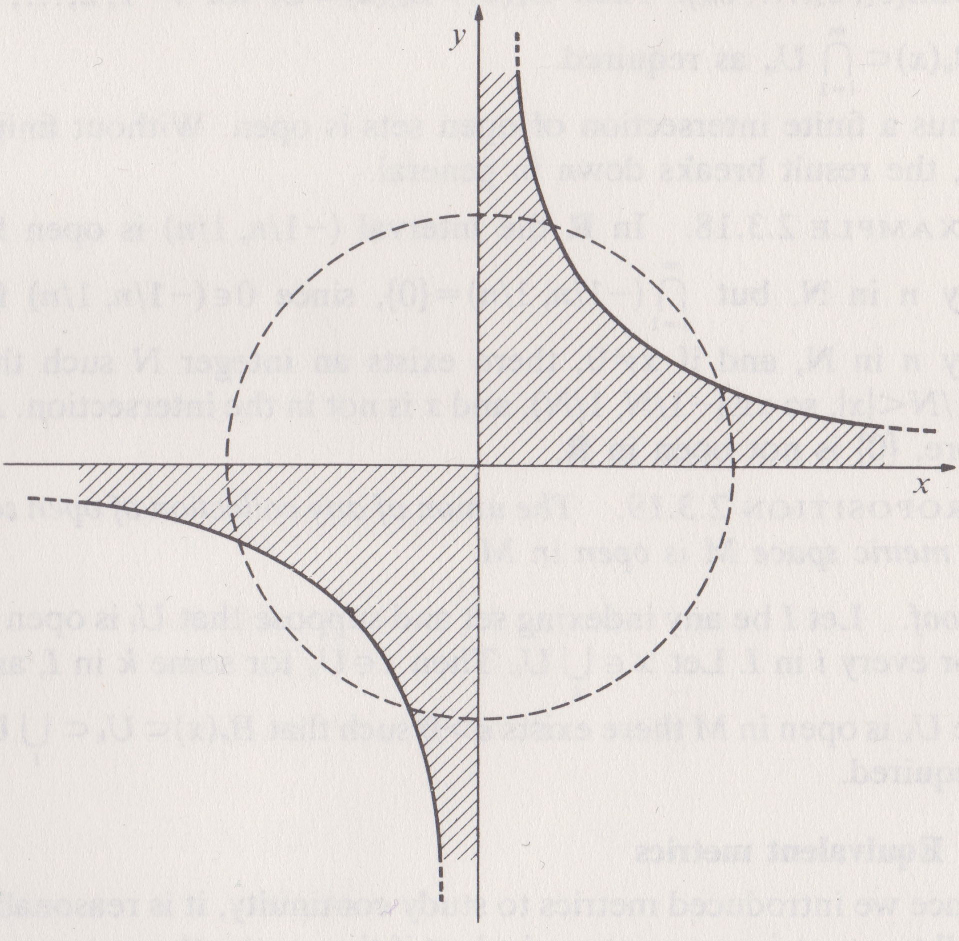


FIG. 2.2

of radius 2 with its centre $(0, 0)$ removed. Hence S, the set of all (x, y) satisfying both conditions, is the intersection of the shaded region with the outlined disc, not including any of the boundary curves or lines.

We could use Proposition 2.3.13 to prove f, g continuous, but in practice one would do this in a different way, which will be stated in § 3.5.

We conclude this section with some propositions which show that 'open set' is a more flexible concept than 'open ball'. They would be false in general if we substituted 'open ball' for every occurrence of 'open set'.

PROPOSITION 2.3.17. *If $U_1, U_2, \ldots, U_m$ are open in a metric space M, then $\bigcap_{i=1}^{m} U_i$ is open in M.*

Proof. Let $x \in \bigcap_{i=1}^{m} U_i$. For each $i = 1, 2, \ldots, m$, then $x \in U_i$ and U_i is open in M, so there exists $\varepsilon_i > 0$ such that $B_{\varepsilon_i}(x) \subset U_i$. Let $\varepsilon = \operatorname{Min}\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m\}$. Then $B_\varepsilon(x) \subset B_{\varepsilon_i}(x) \subset U_i$ for $i = 1, 2, \ldots, m$, so $B_\varepsilon(x) \subset \bigcap_{i=1}^{m} U_i$, as required.

Thus a finite intersection of open sets is open. Without finiteness, the result breaks down in general.

EXAMPLE 2.3.18. In $\mathbf{R}$ the interval $(-1/n, 1/n)$ is open for every n in $\mathbf{N}$, but $\bigcap_{n=1}^{\infty} (-1/n, 1/n) = \{0\}$, since $0 \in (-1/n, 1/n)$ for every n in $\mathbf{N}$, and if $x \neq 0$, there exists an integer N such that $0 < 1/N < |x|$, so $x \notin (-1/N, 1/N)$, and x is not in the intersection. As before, $\{0\}$ is not open in $\mathbf{R}$.

PROPOSITION 2.3.19. *The union of any collection of open sets in a metric space M is open in M.*

Proof. Let I be any indexing set and suppose that U_i is open in M for every i in I. Let $x \in \bigcup_I U_i$. Then $x \in U_k$ for some k in I, and since U_k is open in M there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subset U_k \subset \bigcup_I U_i$, as required.

2.4 Equivalent metrics

Since we introduced metrics to study continuity, it is reasonable to call two metrics on a set equivalent if they make the same maps into and out of that set continuous. The next definition makes this precise. Although the word *topology* itself will not be defined until the next chapter, the name *topologically equivalent* will be used here to avoid confusion with other equivalence relations on metrics which will occur later.

DEFINITION 2.4.1. Two metrics d_1, d_2 on a set A are *topologically equivalent* if for any metric spaces $\{B, d\}$, $\{C, d'\}$ and any maps $f: B \to A$, $g: A \to C$, the following hold:

(a) f is (d, d_1)-continuous iff f is (d, d_2)-continuous,
(b) g is (d_1, d')-continuous iff g is (d_2, d')-continuous.

This is clearly an equivalence relation. The definition in this form is unwieldy, and we next show that it is as good to say that d_1 and d_2 should make the same subsets open.

PROPOSITION 2.4.2. *The following is a necessary and sufficient condition for metrics d_1, d_2 on a set A to be topologically equivalent: a subset U of A is d_1-open iff it is d_2-open.*

Proof. The main point of the proof is that in order to decide which maps into and out of a metric space are continuous, it is enough to look at open sets. Proposition 2.3.13 says this more precisely, and proves sufficiency.

The proof of necessity involves mainly manipulation of definitions. Suppose that d_1 and d_2 are topologically equivalent, and that $U \subset A$ is d_2-open. Let $f: A \to A$ denote the identity map, $f(a) = a$ for every a in A. This is clearly (d_2, d_2)-continuous, hence by Definition 2.4.1(b) it is (d_1, d_2)-continuous. By Proposition 2.3.13 this implies that $f^{-1}(U)$ is d_1-open. But $f^{-1}(U) = U$, so U is d_1-open.

A similar argument, still using Definition 2.4.1(b), shows that if U is d_1-open then it is d_2-open.

We note that it would have been enough to insist on one of (a), (b) in Definition 2.4.1. The point of that definition was not economy but plausibility. We henceforth use the criterion of Proposition 2.4.2 as the definition of topological equivalence of metrics. Before giving examples, we look at a stronger condition implying topological equivalence. There does not appear to be a standard name for this; the name we use is reasonably appropriate, as will be seen in Chapter 9.

PROPOSAL 2.4.3. Two metrics d, d' on a set A will be called *Lipschitz equivalent* if there exist strictly positive constants h, k

such that for any x, y in A,

$$hd'(x, y) \leqslant d(x, y) \leqslant kd'(x, y).$$

This is again clearly an equivalence relation.

PROPOSITION 2.4.4 *Lipschitz equivalent metrics are topologically equivalent.*

Proof. Let d, d' satisfy the above definition of Lipschitz equivalence. It is easy to check the following inclusions:

$$B_{h\varepsilon}(x; d) \subset B_\varepsilon(x; d'), \qquad B_{\varepsilon/k}(x; d') \subset B_\varepsilon(x; d).$$

(Warning: it is also easy to get these the wrong way round.) For example,

$$y \in B_{h\varepsilon}(x; d) \Rightarrow d(x, y) < h\varepsilon \Rightarrow d'(x, y) \leqslant d(x, y)/h < \varepsilon$$
$$\Rightarrow y \in B_\varepsilon(x; d').$$

Now suppose that U is d-open, and let $x \in U$. Then there exists $\varepsilon > 0$ such that $B_\varepsilon(x; d) \subset U$. Hence $B_{\varepsilon/k}(x; d') \subset B_\varepsilon(x; d) \subset U$. This checks that U is d'-open. The converse is proved similarly.

EXAMPLE 2.4.5. Let $A = \mathbf{R}^n$, and let d_1, d_2, d_∞ be as in Example 2.2.3. Exercise 2.6.18 asks the reader to verify that for all x, y in $\mathbf{R}^n$,

$$d_1(x, y) \geqslant d_2(x, y) \geqslant d_\infty(x, y) \geqslant n^{-\frac{1}{2}} d_2(x, y) \geqslant n^{-1} d_1(x, y).$$

This shows that all three metrics are Lipschitz equivalent and hence topologically equivalent. Thus although the open balls with respect to the three metrics are different, the open sets are the same. In any particular situation, it may be slightly more convenient to work with one of these metrics rather than the others. We now know that different choices will not alter anything that concerns continuity alone. Topological equivalence of the product metrics in Example 2.2.7 is proved similarly. Not all metrics on $\mathbf{R}^n$ are topologically equivalent; we have already seen that the discrete metric makes singleton sets open, so it cannot be equivalent to d_1, d_2, d_∞. However, Exercise 5.10.32 is of interest in this connection.

EXAMPLE 2.4.6. On the set of continuous real-valued functions on $[a, b]$ the sup metric d_∞ and the L^1 metric d_1 are not topologically equivalent, although half of a Lipschitz equivalence holds. To see this, let $f, g \in \mathscr{C}[a, b]$. Then $|f(x)-g(x)| \leqslant d_\infty(f, g)$ for all x in $[a, b]$ by definition of d_∞. Hence by integration theory,

$$d_1(f, g) = \int_a^b |f(x)-g(x)|\, dx \leqslant (b-a)\, d_\infty(f, g),$$

which implies that $B_\varepsilon(f; d_\infty) \subset B_{(b-a)\varepsilon}(f; d_1)$.

However, if we let 0 denote also the constant function with value 0, then $B_1(0; d_\infty)$ is not d_1-open. For if it were then we should have $B_\varepsilon(0; d_1) \subset B_1(0; d_\infty)$ for some $\varepsilon > 0$. But for any $\varepsilon > 0$ there exists a continuous function f on $[a, b]$ such that $d_1(f, 0) < \varepsilon$ yet $d_\infty(f, 0) = 1$, so $f \in B_\varepsilon(0; d_1)$ but $f \notin B_1(0; d_\infty)$. Any reader who has difficulty in thinking of such an f is invited to a preview of Fig. 8.2(c) on p. 118.

Closely related to equivalence of metrics are the following definitions, for which we suppose given two distinct metric spaces $M = \{A, d\}$ and $M' = \{A', d'\}$.

DEFINITION 2.4.7. A *topological equivalence* or *homeomorphism* $f: M \rightarrow M'$ is a one–one correspondence such that both f and f^{-1} are continuous.

DEFINITION 2.4.8. A *Lipschitz equivalence* $f: M \rightarrow M'$ is a one–one correspondence such that there exist strictly positive constants h, k satisfying

$$hd'(f(x), f(y)) \leqslant d(x, y) \leqslant kd'(f(x), f(y))$$

for all x, y in M.

As before, a Lipschitz equivalence is a topological equivalence. A special case of Lipschitz equivalence is given in the next definition.

DEFINITION 2.4.9. An *isometry* $f: M \rightarrow M'$ is a one–one correspondence such that $d'(f(x), f(y)) = d(x, y)$ for all x, y in M.

We also use the term *isometry into* for a map which satisfies the conditions for an isometry except that it is not necessarily onto M'.

EXAMPLE 2.4.10. Let $f: \mathbf{R}^2 \rightarrow \mathbf{C}$ be given by $f(x_1, x_2) = x_1 + ix_2$. Then f is an isometry from $\mathbf{R}^2$ with the Euclidean metric to $\mathbf{C}$ with the metric of Example 2.2.4.

The relation between Definitions 2.4.1, 2.4.3 and Definitions 2.4.7, 2.4.8 is as follows. Two metrics d, d' on a set A are topologically (Lipschitz) equivalent iff the identity map of A is a topological (Lipschitz) equivalence from $\{A, d\}$ to $\{A, d'\}$. Conversely, if $f: M \to M'$ is a topological (Lipschitz) equivalence then the formula $d''(x, y) = d'(f(x), f(y))$ gives a metric on A which is topologically (Lipschitz) equivalent to d. The proofs are easy.

2.5 Continuity

In this chapter we began by looking at metric spaces as a generalized framework in which to study continuity. However, §§ 2.3 and 2.4 suggest two points. First, it is not so much the particular metrics on spaces M, M' which decide whether a map from M to M' is continuous, as the topological equivalence class of the metrics, or in other words the 'open set structure' of M and M' (against this, it may be argued that in any practical application, it is likely that one will check continuity by means of particular metrics). Secondly, in order to define continuity of a map from one set to another, perhaps after all we do not really need metrics on the sets, but only some adequate notion of 'open subsets'. What properties these open sets should have is again a matter of historical trial and error. It has been found that the theory of continuity proceeds very pleasantly if we just abstract those properties of open sets given in Example 2.3.10 and Propositions 2.3.17 and 2.3.19. This is the subject of the next chapter.

Exercises 2.6

In the first six exercises, $M = \{A, d\}$ is an arbitrary metric space.

1. Prove that if x, y are two points in M with $d(x, y) = 2r > 0$, then $B_r(x)$ and $B_r(y)$ are disjoint.
2. Given any three points x, y, z in M, prove that
$$|d(x, z) - d(y, z)| \leqslant d(x, y).$$
3. Given any four points x, y, z, t in M, prove that
$$d(x, z) + d(y, t) \geqslant |d(x, y) - d(z, t)|.$$
4. Prove that if $C \subset B$ where B is a bounded subset of M, then C is bounded and $\operatorname{diam} C \leqslant \operatorname{diam} B$.
5. Prove that if B, C are bounded subsets of M with $B \cap C \neq \varnothing$ then $\operatorname{diam}(B \cup C) \leqslant \operatorname{diam} B + \operatorname{diam} C$.

6*. For any x, y in A, let

$$d_1(x, y) = kd(x, y) \qquad \text{for some constant } k>0,$$
$$d_2(x, y) = \mathrm{Min}\{1, d(x, y)\},$$
$$d_3(x, y) = d(x, y)/(1+d(x, y)),$$
$$d_4(x, y) = (d(x, y))^2.$$

Prove that d_1, d_2, and d_3 are metrics for A, but that d_4 may or may not be a metric for A.

7. Think of physical situations in which d_1, d_∞ of Example 2.2.3 would be more appropriate than d_2. (Hint: in the U.S.A. d_1 is sometimes called the taxicab metric.)
8. Let A be the circle $\{(x_1, x_2) \in \mathbf{R}^2 : x_1^2 + x_2^2 = 1\}$, and for x, y in A let $d(x, y) = 0$ if $x = y$, $d(x, y) = \pi$ if x, y are antipodal, and otherwise let $d(x, y)$ be the length of the shorter arc of A joining x and y. Prove that d is a metric for A.
9. Draw pictures of $B_1(0; d_1)$ and $B_1(0; d_\infty)$, where $0 = (0, 0) \in \mathbf{R}^2$ and d_1, d_∞ are as in Example 2.2.3.
10. Prove that a subset of a metric space is open iff it is a union of open balls.
11. Prove that the rectangle $(a, b) \times (c, d)$ is open in $\mathbf{R}^2$.
12. Is it possible in a metric space M containing more than one point that the only open subsets are M and $\varnothing$? (Hint: use Exercise 1.)
13. Suppose that f is a continuous real-valued function on a metric space M and that $f(a)>0$ for some a in M. Show that there exists some $\delta>0$ such that $f(x)>0$ for all x in $B_\delta(a)$.
14. Given that f, g are continuous real-valued functions on a metric space M, prove that $f+g$, $f-g$, λf, $|f|$, $\mathrm{Max}\{f, g\}$, $\mathrm{Min}\{f, g\}$, $f.g$ defined as in Exercises 1.5 are continuous on M, and that f/g is continuous at any a in M such that $g(a) \neq 0$.
15. Show that if M, M' are metric spaces and $M \times M'$ has one of the metrics d_1, d_2, d_∞ of Example 2.2.7, then

$$p_1 : M \times M' \to M, \qquad p_2 : M \times M' \to M'$$

given by $p_1(x, y) = x$, $p_2(x, y) = y$ are continuous.

16. Let $M \times M'$ be as in Exercise 15, and for some fixed a, b in M, M' respectively let $i_b : M \to M \times M'$ and $i_a : M' \to M \times M'$ be

given by $i_b(x)=(x, b)$, $i_a(y)=(a, y)$. Prove that i_b and i_a are continuous.

17. Given a metric space $M=\{A, d\}$, let $M\times M=\{A\times A, d_1\}$ be as in Example 2.2.7. Show that $d: A\times A\to\mathbf{R}$ is a continuous function from $M\times M$ to $\mathbf{R}$. (Hint: use Exercise 3.)
18. Verify the inequalities in Example 2.4.5.
19. Prove that the metric d in Exercise 8 is Lipschitz equivalent to the metric on the circle induced by the Euclidean metric on $\mathbf{R}^2$.
20*. Prove that the metrics d_2, d_3 in Exercise 6 are topologically equivalent to the metric d from which they are constructed.
21. A translation $T:\mathbf{R}^2\to\mathbf{R}^2$ is a map given by $T(x_1, x_2)=(x_1+a, x_2+b)$ for some fixed (a, b) in $\mathbf{R}^2$, and a rotation $\rho:\mathbf{R}^2\to\mathbf{R}^2$ is a map given by

$$\rho(x_1, x_2)=(x_1\cos\theta-x_2\sin\theta,\ x_1\sin\theta+x_2\cos\theta)$$

for some fixed θ in $\mathbf{R}$. Prove that the Euclidean metric d_2 on $\mathbf{R}^2$ is translation- and rotation-invariant, in the sense that for any (x, y) in $\mathbf{R}^2$ and any translation T, rotation ρ,

$$d_2(T(x), T(y))=d_2(x, y)=d_2(\rho(x), \rho(y)).$$

22. Are d_1 and/or d_∞ of Example 2.2.3 translation- and/or rotation-invariant?
23. Let p be a fixed prime number ($p>0$). Define $d:\mathbf{Z}\times\mathbf{Z}\to\mathbf{R}$ by $d(m, m)=0$, $d(m, n)=1/r$ where $m-n=p^{r-1}k$ (r, k integers, k not divisible by p). Prove that d is a metric on $\mathbf{Z}$.
24*. Construct a function $f:\mathbf{R}^2\to\mathbf{R}$ such that f is not continuous at $0=(0, 0)$, but for every straight line L through 0 in $\mathbf{R}^2$, $f\mid L$ is continuous at 0.
25*. Let $\mathscr{C}=\mathscr{C}[a, b]$ be as in Example 2.2.8. Suppose that $f, g\in\mathscr{C}$ with $f(x)<g(x)$ for all x in $[a, b]$. Is the set $\{h\in\mathscr{C}: f(x)<h(x)<g(x)$ for all x in $[a, b]\}$ an open ball in $\mathscr{C}$?
26. Suppose that in a metric space M, $B_r(x)=B_s(y)$ for some x, y in M and r, s in $\mathbf{R}$. Is $x=y$? Is $r=s$?
27. How big are the bumps in the road surface in Example 2.2.17? Could you afford to pay for a trip of 1 km measured along the x-axis if the charge is 1 p per 1000 km measured along the graph of g?

3

Continuity generalized: topological spaces

3.1 Topological spaces

THE framework for studying continuity hinted at in § 2.5 is that of topological spaces. In view of the remarks made there, the following is a plausible definition.

DEFINITION 3.1.1. A *topological space* $T = \{A, \mathcal{T}\}$ consists of a non-empty set A together with a fixed collection $\mathcal{T}$ of subsets of A satisfying

(T1) $A, \emptyset \in \mathcal{T}$,

(T2) the intersection of any two sets in $\mathcal{T}$ is again in $\mathcal{T}$,

(T3) the union of any collection of sets in $\mathcal{T}$ is again in $\mathcal{T}$.

The collection $\mathcal{T}$ is called a *topology* for A, and the members of $\mathcal{T}$ are called the *open sets* of T. Thus '$U \in \mathcal{T}$' and 'U is open in T' mean the same thing. Elements of A are called points in the space T. In practice we often use the same name for A and T when it is unmistakable which topology is intended. It is important to remember that $\mathcal{T}$ is in general only a *sub*collection of the set of all subsets of A.

From (T2) it follows by induction that the intersection of any finite collection of sets in $\mathcal{T}$ is again in $\mathcal{T}$.

Bearing in mind Proposition 2.3.13 we define continuity as follows.

DEFINITION 3.1.2. Given two topological spaces $T_1 = \{A_1, \mathcal{T}_1\}$, $T_2 = \{A_2, \mathcal{T}_2\}$ and a map $f: A_1 \rightarrow A_2$, we say that f is *continuous* (with respect to the topologies $\mathcal{T}_1$, $\mathcal{T}_2$) if $U \in \mathcal{T}_2 \Rightarrow f^{-1}(U) \in \mathcal{T}_1$. If necessary for clarity we say that f is $(\mathcal{T}_1, \mathcal{T}_2)$-continuous.

It is important to note that this definition concerns *inverse* images of open sets in T_2 and makes no assertion about (direct) images of open sets in T_1 (cf. Warning 2.3.14).

On previous occasions we defined continuity at a point first. We may still do that, although at the present level of generality, Definition 3.1.2 is the one most used.

DEFINITION 3.1.3. With the notation of Definition 3.1.2, let $a \in A_1$. Then f is *continuous at a* if given any U_2 in $\mathcal{T}_2$ such that $f(a) \in U_2$, there exists U_1 in $\mathcal{T}_1$ such that $a \in U_1$ and $f(U_1) \subset U_2$.

Then (Exercise 3.9.1) f is continuous iff f is continuous at every point in A_1.

Our first example of a topological space is a pathological one. This will facilitate the discussion of the second example.

EXAMPLE 3.1.4. *Discrete spaces.* Let A be any non-empty set and let $\mathcal{T}$ be the collection of all subsets of A, $\mathcal{T} = 2^A$. The topological space axioms are obviously satisfied by this $\mathcal{T}$. We call $\mathcal{T}$ the discrete topology on A and $\{A, \mathcal{T}\}$ a discrete space.

EXAMPLE 3.1.5. Any metric space $M = \{A, d\}$ gives rise to a topological space $T = \{A, \mathcal{T}\}$, where $\mathcal{T}$ is defined to be the collection of all those subsets which are open in the sense of Definition 2.3.8. By Example 2.3.10 and Propositions 2.3.17 and 2.3.19, the topological space axioms are satisfied by this collection.

This use of the term *open* in two slightly different contexts is confusing at first. On the one hand, when a metric space $M = \{A, d\}$ has been specified, we can work out, using Definition 2.3.8, whether or not a given subset of A is d-open. For example, given a discrete metric space $\{A, d\}$ we proved (Example 2.3.11) that any subset of A is d-open. On the other hand, when a topological space $T = \{A, \mathcal{T}\}$ has been specified, then to see whether a given subset of A is open in T (that is, belongs to $\mathcal{T}$), we have only to 'look' at the list of subsets in $\mathcal{T}$, there is nothing to prove. For example in a discrete topological space (Example 3.1.4) any subset is open because of the specification of the discrete topology. Example 3.1.5 says that given a metric space $M = \{A, d\}$, we may construct a topological space $\{A, \mathcal{T}\}$ by defining $\mathcal{T}$ to consist of exactly those subsets of A which are d-open. A topological space which arises in this way from a metric space is called metrizable. Our discussion of topologically equivalent metrics in § 2.4 shows that distinct metric spaces may give rise to the same topological space. For example, the metrics d_1, d_2, d_∞ on $\mathbf{R}^n$ all give rise to the same topology. Whenever we refer to $\mathbf{R}^n$ as a topological space, this topology will be understood unless some other is specified.

A separate point is that a given set may admit more than one topology. For example, the discrete topology on $\mathbf{R}^n$ is distinct from the usual topology, since singleton sets are not open in the usual topology. If $\mathcal{T}_1$ and $\mathcal{T}_2$ are topologies for the same set, we say that $\mathcal{T}_1$ is finer (coarser) than $\mathcal{T}_2$ if $\mathcal{T}_1 \supset \mathcal{T}_2$ $(\mathcal{T}_1 \subset \mathcal{T}_2)$. The term weaker has also been used (indeed, more often) in this connection, but unfortunately with opposite meanings on different occasions. Going along with majority opinion, we take weaker to mean coarser. Given two topologies on a set, it may be that neither is weaker than the other (see Exercise 3.9.3). The discrete topology on a set A is finer than any other topology on A. At the other extreme, there is another pathological example.

EXAMPLE 3.1.6. Take A to be any non-empty set and let $\mathcal{T} = \{A, \emptyset\}$. Then it is easy to check that the axioms for a topology hold. This is called the indiscrete topology on A. It is weaker than any other topology on A, since by (T1) any topology on A contains the indiscrete topology. When A contains more than one point, this topology is not metrizable. For suppose that d is any metric on A. Let x, y be distinct points in A. Then $\varepsilon = d(x, y) > 0$, so $B_{\frac{1}{2}\varepsilon}(x)$ is a d-open set, and hence belongs to the topology to which d gives rise. But $x \in B_{\frac{1}{2}\varepsilon}(x)$ while $y \notin B_{\frac{1}{2}\varepsilon}(x)$, so $B_{\frac{1}{2}\varepsilon}(x)$ is neither $\emptyset$ nor A. Hence the topology to which d gives rise is not the indiscrete topology. The reason for this will be formalized in Chapter 4.

EXAMPLE 3.1.7. Let $A = \mathbf{R}$ and let $\mathcal{T}$ consist of all subsets U of $\mathbf{R}$ such that $\mathbf{R} - U$ is finite, together with $\emptyset$. Clearly (T1) holds. To prove (T2) note that if $\mathbf{R} - U$ and $\mathbf{R} - V$ are finite then so is $\mathbf{R} - (U \cap V) = (\mathbf{R} - U) \cup (\mathbf{R} - V)$, while if $U = \emptyset$ or $V = \emptyset$ then $U \cap V = \emptyset$. To prove (T3), suppose that for each i in some set I, either $U_i = \emptyset$ or $\mathbf{R} - U_i$ is finite. Then either $U_i = \emptyset$ for all i in I and $\bigcup_I U_i = \emptyset$, or $\mathbf{R} - U_j$ is finite for some j in I and $\mathbf{R} - \bigcup_I U_i = \bigcap_I (\mathbf{R} - U_i)$ is finite.

Example 3.1.6 shows that the notion of topological space is strictly more general than that of metric space. However, it does not show this in a very interesting way. There *are* interesting examples of non-metrizable topological spaces (notably in algebraic geometry) and there are also situations (for example, in geometric and differential topology) in which although the spaces are metrizable, it is rather irrelevant to use metrics. Example 3.1.7 is a very simple example of the Zariski topologies used in algebraic geometry; an example in which metrics are, at least for many purposes, irrelevant will be given in § 3.8.

We shall not prove much before introducing further hypotheses about topological spaces, but in the rest of this chapter we introduce some constructions and terminology. The next proposition illustrates the simplicity and triviality of results available for topological spaces unqualified by extra assumptions.

PROPOSITION 3.1.8. *If* T_1, T_2, T_3 *are topological spaces and* $f: T_1 \to T_2$, $g: T_2 \to T_3$ *are continuous maps, then* $g \circ f: T_1 \to T_3$ *is continuous.*

Proof. Let U be any open subset of T_3. Then $g^{-1}(U)$ is open in T_2 by continuity of g, so $f^{-1}(g^{-1}(U))$ is open in T_1 by continuity of f. But $f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U)$. Hence $g \circ f$ is continuous.

This generalizes propositions asserting continuity of a 'function of a function' in several particular cases.

EXAMPLE 3.1.9. If f, g are real-valued functions of a real variable such that f is continuous on $[a, b]$ and g is continuous on some subspace of $\mathbf{R}$ containing $f([a, b])$, then $g \circ f$ is continuous on $[a, b]$.

The proof of the next proposition is Exercise 3.9.5.

PROPOSITION 3.1.10. (a) *The identity map of any topological space is continuous.*

(b) *Any constant map is continuous.*

PROPOSITION 3.1.11. *Suppose that, for* $i = 1, 2$, *the metric space* $M_i = \{A_i, d_i\}$ *gives rise to the topological space* $T_i = \{A_i, \mathcal{T}_i\}$ *and that* $f: A_1 \to A_2$ *is* (d_1, d_2)*-continuous. Then* f *is* $(\mathcal{T}_1, \mathcal{T}_2)$*-continuous.*

This follows from Proposition 2.3.13 and Definition 3.1.2.

†The process of going from a metric space to the topological space it gives rise to, and the corresponding process (see Proposition 3.1.11) for continuous maps, make up a covariant functor from the category of metric spaces and continuous maps to the category of topological spaces and continuous maps. This is a convenient, precise description, whose truth will be seen by readers who know the language of categories and functors. Those who do not are promised that this is no barrier to reading further.†

3.2 Bases

In a metric space, any open set is a union of open balls (Exercise 2.6.10) where in general infinitely many balls are involved in the union. In a topological space it is often convenient to have some subcollection of the open sets, called a basis, having this same property as the open balls in a metric space. Bases have two aspects, which we might call the analytic and the synthetic. So far we have emphasized the analytic aspect, whose definition we now formalize.

DEFINITION 3.2.1. Given a topological space $T=\{A, \mathcal{T}\}$, a *basis* for $\mathcal{T}$ is a subcollection $\mathcal{B}\subset\mathcal{T}$ such that every set in $\mathcal{T}$ is a union of sets from $\mathcal{B}$.

The synthetic aspect of bases is as follows.

DEFINITION 3.2.2. Given a set A, a collection $\mathcal{B}$ of subsets of A is a (*synthetic*) *basis* for A if

(B1) A is a union of sets from $\mathcal{B}$,

(B2) if $B_1, B_2\in\mathcal{B}$ then $B_1\cap B_2$ is a union of sets from $\mathcal{B}$.

The point of this definition is explained by the next proposition.

PROPOSITION 3.2.3 *Suppose that $\mathcal{B}$ is a basis for a non-empty set A, and let $\mathcal{T}=\{U\in 2^A : U$ is a union of sets from $\mathcal{B}\}$. Then $\mathcal{T}$ is a topology for A.*

Proof. (T1) By (B1), $A\in\mathcal{T}$. In the statement of Proposition 3.2.3 it is understood that we are allowed to take the union of no sets from $\mathcal{B}$, so $\emptyset\in\mathcal{T}$.

(T2) If $U=\bigcup_{i\in I} B_{i1}$, $V=\bigcup_{j\in J} B_{j2}$ for some indexing sets I, J, where the B_{i1} and the B_{j2} all belong to $\mathcal{B}$, then

$$U\cap V=\bigcup_{(i,j)\in I\times J}(B_{i1}\cap B_{j2}),$$

which is a union of unions of sets from $\mathcal{B}$ by (B2), hence a union of sets from $\mathcal{B}$.

(T3) As we have just noted, a union of unions of sets from $\mathcal{B}$ is again a union of sets from $\mathcal{B}$.

The two kinds of bases are compatible in the sense of the next proposition, which follows from the definitions.

PROPOSITION 3.2.4. *If $\mathcal{T}$ is the topology on a set A arising from a (synthetic) basis $\mathcal{B}$ for A, then $\mathcal{B}$ is a basis for $\mathcal{T}$. If $\{A, \mathcal{T}\}$ is* *a topological space and $\mathscr{B}$ is a basis for $\mathscr{T}$ then $\mathscr{B}$ is a (synthetic) basis for A and the topology it gives rise to is $\mathscr{T}$.*

The two aspects of bases correspond to two ways in which they are used.

APPLICATION 3.2.5. To check that $f: T_1 \to T_2$ is continuous, it is enough to prove that, for each set B in some basis $\mathscr{B}$ for the topology of T_2, $f^{-1}(B)$ is open in T_1. For suppose that this has been proved. Then any open set U in T_2 is of the form $\bigcup_I B_i$ where $B_i \in \mathscr{B}$ for each i in some indexing set I. Hence

$$f^{-1}(U)=f^{-1}\left(\bigcup_I B_i\right)=\bigcup_I (f^{-1}B_i).$$

This last set is a union of sets known to be open in T_1, hence $f^{-1}(U)$ is open in T_1.

APPLICATION 3.2.6. As a way of specifying a topology on a set, we may say 'Let $\mathscr{T}$ be the topology with the sets... as basis'. In this case, the named sets should satisfy (B1) and (B2).

A given topology may have many distinct bases. Particularly for Application 3.2.5, it is desirable to have a basis with as 'few' sets in it as possible.

EXAMPLE 3.2.7. We began this section by noting that in any metric space the open balls form a basis for the topology arising from the metric. We can often get a more economical basis than this. For example, let $S=\{(x, y)\in \mathbf{R}^2 : x, y \in \mathbf{Q}\}$, and let $\mathscr{B}$ denote the collection $\{B_q(s)\}$ as s ranges over S and q ranges over $\mathbf{Q}_+$. Exercise 3.9.8 asks the reader to show that $\mathscr{B}$ is a countable basis for the usual topology on $\mathbf{R}^2$. A topological space which admits a countable basis is called second countable. We shall not study second countability in this book.

*3.3 Sub-bases and weak topologies

Sub-bases are similar to, but more general than, bases. We describe them briefly.

A sub-basis for a topology $\mathscr{T}$ is a subcollection $\mathscr{S} \subset \mathscr{T}$ such that any set in $\mathscr{T}$ is a union of finite intersections of sets from $\mathscr{S}$. Thus a sub-basis tends to be even more economical than a basis.

EXAMPLE 3.3.1. In $\mathbf{R}$, the collection of all open intervals is a basis for the usual topology. For a sub-basis, it is enough to take the open intervals of the form $(-\infty, b)$ and (a, ∞), since any open interval is either one of these or else the intersection of two of them, $(a, b)=(-\infty, b)\cap(a, \infty)$; and any open set in $\mathbf{R}$ is a union of open intervals.

The analogue of Application 3.2.5 holds with sub-basis replacing basis.

Synthetically, a sub-basis $\mathcal{S}$ for a set A is any collection of subsets of A. The collection of all finite intersections of sets from $\mathcal{S}$ (including A as the intersection of no sets from $\mathcal{S}$) then forms a synthetic basis for A, so the collection of all unions of finite intersections of sets from $\mathcal{S}$ gives a topology (unless $A=\varnothing$, which is excluded in Definition 3.1.1). This is called the topology generated by $\mathcal{S}$; it may also be described as the weakest topology containing $\mathcal{S}$, or the intersection of all topologies containing $\mathcal{S}$. As in Application 3.2.6, we may specify a topology by saying 'Let $\mathcal{T}$ be the topology with the sets... as sub-basis'. In this case, there are no conditions to be checked for the named sets.

A related idea is that of weak topologies. Suppose that we are given a set A and an I-indexed family of maps $f_i : A \rightarrow A_i$, where $T_i=\{A_i, \mathcal{T}_i\}$ is a topological space for each i in I. The weak topology on A with respect to $\{f_i : i \in I\}$ is the weakest topology $\mathcal{T}$ on A such that each f_i is $(\mathcal{T}, \mathcal{T}_i)$-continuous. It coincides with the topology generated by the sub-basis $\{f_i^{-1}(U_i) : U_i \in \mathcal{T}_i, i \in I\}$. This way of specifying a topology is often used in functional analysis.*

3.4 Subspaces

As for metric spaces, we may form subspaces of topological spaces.

DEFINITION 3.4.1. Given a topological space $T=\{A, \mathcal{T}\}$ and a non-empty subset $H\subset A$, *the relative, induced,* or *subspace* topology on H is $\mathcal{T}_H=\{U\cap H : U \in \mathcal{T}\}$.

It is easily checked that $\{H, \mathcal{T}_H\}$ is a topological space. It is called a subspace of T. Since we loosely use the same name for a topological space and its underlying set, H is referred to as a subset or as a subspace; in the latter case, it is always understood to have the subspace topology.

Metric and topological subspaces are compatible constructions in the sense of Exercise 3.9.13. The definition of the subspace topology is fairly simple, and has been given above without motivation. We shall now 'rediscover' it as the only topology on H satisfying certain desirable properties. Let $i: H \to A$ denote the inclusion map. Let $T' = \{A', \mathcal{T}'\}$ be any topological space and let $f: A \to A'$, $g: A' \to H$ be maps. We note that $f \mid H = f \circ i$.

PROPOSITION 3.4.2. *With the above notation:*

(a) *if* f *is continuous then so is* $f \circ i$;

(b) g *is continuous iff* $i \circ g$ *is continuous.*

Proof. We begin with a criticism of the statement. In order to emphasize its simplicity we have cut down on notation. A more precise version is:

(a) if f is $(\mathcal{T}, \mathcal{T}')$-continuous then $f \circ i$ is $(\mathcal{T}_H, \mathcal{T}')$-*continuous,*

(b) g is $(\mathcal{T}', \mathcal{T}_H)$-continuous iff $i \circ g$ is $(\mathcal{T}', \mathcal{T})$-continuous.

To prove (a) it is enough by Proposition 3.1.8 to prove that i is $(\mathcal{T}_H, \mathcal{T})$-continuous. This follows since if $U \in \mathcal{T}$ then

$$i^{-1}(U) = U \cap H \in \mathcal{T}_H.$$

The reader may conveniently follow the rest of the proof using the diagram:

$$\begin{tikzcd}
\{A', \mathcal{T}'\} \arrow[rr, "g"] \arrow[dr, "i \circ g"'] & & \{H, \mathcal{T}_H\} \arrow[dl, "i"] \\
& \{A, \mathcal{T}\} &
\end{tikzcd}$$

If g is continuous then so is $i \circ g$ as in (a). Suppose that $i \circ g$ is continuous and let $V \in \mathcal{T}_H$. Then $V = U \cap H$ for some U in $\mathcal{T}$, by definition of $\mathcal{T}_H$, and $U \cap H = i^{-1}(U)$, so $g^{-1}(V) = g^{-1}(i^{-1}(U)) = (i \circ g)^{-1}(U)$. This last set is in $\mathcal{T}'$ by continuity of $i \circ g$. Hence g is continuous.

*PROPOSITION 3.4.3. *With the above notation, the induced topology* $\mathcal{T}_H$ *is the only topology on* H *which satisfies Proposition 3.4.2(b) for all possible* g.

Proof. The reader is encouraged to draw diagrams similar to the one in the previous proof while reading this proof. Suppose that $\mathcal{T}''$ is a topology on H such that for any topological space $\{A', \mathcal{T}'\}$ and any map $g: A' \to H$, g is $(\mathcal{T}', \mathcal{T}'')$-continuous iff $i \circ g$ is $(\mathcal{T}', \mathcal{T})$-continuous.

First take $A' = H$, $\mathcal{T}' = \mathcal{T}''$, and g the identity map of H. Since g is certainly $(\mathcal{T}'', \mathcal{T}'')$-continuous, $i\circ g$ is $(\mathcal{T}'', \mathcal{T})$-continuous. Hence for any U in $\mathcal{T}$, $(i\circ g)^{-1}(U)\in\mathcal{T}''$. But $(i\circ g)^{-1}(U)=i^{-1}(U)=U\cap H$. Hence $\mathcal{T}_H\subset\mathcal{T}''$.

Secondly take $A'=H$, $\mathcal{T}'=\mathcal{T}_H$, and g the identity map of H. Since $i\circ g=i$ is $(\mathcal{T}_H, \mathcal{T})$-continuous, g is $(\mathcal{T}_H, \mathcal{T}'')$-continuous. But by definition of continuity, this is the same as saying $\mathcal{T}''\subset\mathcal{T}_H$. Hence $\mathcal{T}''=\mathcal{T}_H$, as required.*

3.5 Products

When we write parametric equations for a curve in the plane, say,

$$\left.\begin{array}{l} x=x(t)\\ y=y(t)\end{array}\right\}\quad a\leqslant t\leqslant b,$$

we expect the curve to be continuous (as a function from $[a, b]$ to $\mathbf{R}^2$) provided the 'coordinate functions' $x(t)$ and $y(t)$ are continuous (as functions from $[a, b]$ to $\mathbf{R}$). In elementary mathematics this is taken for granted, or else the curve is defined to be continuous provided x and y are continuous functions of t. Later, continuity of a function $f:[a, b]\to\mathbf{R}^2$ is defined in terms of the Euclidean metrics on $[a, b]$ and $\mathbf{R}^2$, and it is then an easy proposition that f is continuous iff the first and second coordinate functions are continuous. These coordinate functions x and y are the compositions $p_1\circ f$ and $p_2\circ f$, where p_1, p_2 are the projections of $\mathbf{R}^2$ onto the x-axis and the y-axis respectively.

More generally, for any two sets A_1, A_2 we define projections $p_1: A_1\times A_2\to A_1$, $p_2: A_1\times A_2\to A_2$ by $p_1(x, y)=x$, $p_2(x, y)=y$, for any (x, y) in $A_1\times A_2$. A map $f: A'\to A_1\times A_2$ from any set A' then determines corresponding coordinate maps $p_1\circ f: A'\to A_1$ and $p_2\circ f: A'\to A_2$. Conversely any two maps $g: A'\to A_1$, $h: A'\to A_2$ give rise to a map $f: A'\to A_1\times A_2$ such that $p_1\circ f=g$, $p_2\circ f=h$, where $f(z)=(g(z), h(z))$ for any z in A'.

Now suppose that $T_1=\{A_1, \mathcal{T}_1\}$, $T_2=\{A_2, \mathcal{T}_2\}$ are topological spaces. By analogy with the plane, we want to put a topology $\mathcal{T}$ on $A_1\times A_2$ so that for any topological space $T'=\{A', \mathcal{T}'\}$ and any map $f: A'\to A_1\times A_2$, f is continuous iff $p_1\circ f$ and $p_2\circ f$ are continuous.

*Suppose that this is possible, and that $\mathcal{T}$ is such a topology. Take $T'=\{A_1\times A_2, \mathcal{T}\}$ and f to be the identity map of $A_1\times A_2$. Then $p_1\circ f=p_1$, $p_2\circ f=p_2$, so since the identity map is continuous, p_1 and p_2 must be continuous. Let $U_1\in\mathcal{T}_1$. By continuity of p_1, $p_1^{-1}(U_1)=U_1\times A_2$ is in $\mathcal{T}$. Similarly $A_1\times U_2$ is in $\mathcal{T}$ for any U_2 in $\mathcal{T}_2$. Hence by

(T2), $U_1 \times U_2 = (U_1 \times A_2) \cap (A_1 \times U_2) \in \mathcal{T}$, and by (T3) it follows that $\mathcal{T}'' \subset \mathcal{T}$, where $\mathcal{T}''$ is the collection of all unions of sets of the form $U_1 \times U_2$ for U_1 in $\mathcal{T}_1$, U_2 in $\mathcal{T}_2$.

Now suppose that we can prove that $\mathcal{T}''$ is a topology for $A_1 \times A_2$ (as the reader will guess from the loaded notation, we *shall* prove this). Take $T' = \{A_1 \times A_2, \mathcal{T}''\}$ and f to be the identity map of $A_1 \times A_2$. Then again $p_1 \circ f = p_1$, $p_2 \circ f = p_2$, and by our choice of $\mathcal{T}''$, $p_1 : T' \to T_1$ and $p_2 : T' \to T_2$ are continuous. Hence f is $(\mathcal{T}'', \mathcal{T})$-continuous. But by definition of continuity this means that $\mathcal{T} \subset \mathcal{T}''$. Hence $\mathcal{T} = \mathcal{T}''$. The upshot of all this rigmarole is that, provided we can show that $\mathcal{T}''$ is a topology, then it is the unique topology for which our original aim is fulfilled.*

We now make a fresh start. The only loss to readers who skipped the previous two paragraphs will be a sense of the inevitability of the next definition.

DEFINITION 3.5.1. Given topological spaces $T_1 = \{A_1, \mathcal{T}_1\}$, $T_2 = \{A_2, \mathcal{T}_2\}$, the *product topology* $\mathcal{T}$ for $A_1 \times A_2$ is the topology with basis

$$\mathcal{B} = \{U_1 \times U_2 : U_1 \in \mathcal{T}_1, U_2 \in \mathcal{T}_2\}.$$

The space $T_1 \times T_2 = \{A_1 \times A_2, \mathcal{T}\}$ is called the *topological product* of T_1 and T_2.

In order to make the definition respectable, we have to check that the conditions (B1) and (B2) for a basis are satisfied (see Definition 3.2.2). Since $A_1 \in \mathcal{T}_1$, $A_2 \in \mathcal{T}_2$, it follows that $A_1 \times A_2 \in \mathcal{B}$, so (B1) holds. To prove (B2), suppose that $U_1, V_1 \in \mathcal{T}_1$, $U_2, V_2 \in \mathcal{T}_2$. The following set-theoretic equation is easily proved:

$$(U_1 \times U_2) \cap (V_1 \times V_2) = (U_1 \cap V_1) \times (U_2 \cap V_2).$$

Since $U_1 \cap V_1 \in \mathcal{T}_1$ and $U_2 \cap V_2 \in \mathcal{T}_2$, we have

$$(U_1 \cap V_1) \times (U_2 \cap V_2) \in \mathcal{B}.$$

Thus $(U_1 \times U_2) \cap (V_1 \times V_2)$ is a union of sets (one in number) from $\mathcal{B}$, as required.

Metric and topological products are compatible constructions in the sense of Exercise 3.9.14.

Warning. One of the commonest errors in topology is to assume that every open set in $T_1 \times T_2$ is of the form $U_1 \times U_2$ where $U_1 \in \mathcal{T}_1$, $U_2 \in \mathcal{T}_2$. It has to be remembered that these are only the open sets in a certain basis for the product topology. Any open set in $T_1 \times T_2$ is a union of sets of this special form, but in general infinitely many sets are involved in the union. For example if $0=(0, 0)\in \mathbf{R}^2$, then $B_1(0)$ is an open set in $\mathbf{R}^2$ with the product topology, since by Exercise 3.9.14 this coincides with the Euclidean topology. But $B_1(0)$ is not even set-theoretically of the form $U_1 \times U_2$ for subsets U_1, U_2 of $\mathbf{R}$. (This is intuitively obvious. To prove it, suppose that $B_1(0)=U_1 \times U_2$. Then since $(\frac{3}{4}, 0)\in B_1(0)$ we must have $\frac{3}{4}\in U_1$. Similarly $\frac{3}{4}\in U_2$, so $(\frac{3}{4}, \frac{3}{4})\in U_1 \times U_2$. But $(\frac{3}{4}, \frac{3}{4})\notin B_1(0)$.)

The next proposition is an immediate consequence of Definition 3.5.1.

PROPOSITION 3.5.2. *Given a point* (x, y) *in* $T_1 \times T_2$ *and an open set* W *in* $T_1 \times T_2$ *with* $(x, y)\in W$, *there exist sets* U_1, U_2 *open in* T_1, T_2 *respectively such that* $(x, y)\in U_1 \times U_2 \subset W$.

REMARK. When we write a set W in $T_1 \times T_2$ as a union of sets of the form $U_1 \times U_2$, we might as well assume that all the U_1, U_2 involved are non-empty, since $U_1 \times \varnothing = \varnothing \times U_2 = \varnothing$.

The next two propositions were proved in a rather contorted way before Definition 3.5.1. For convenience we reprove them now.

PROPOSITION 3.5.3. *The projections* $p_1 : T_1 \times T_2 \to T_1$ *and* $p_2 : T_1 \times T_2 \to T_2$ *of a topological product onto its factors are continuous.*

Proof. If U is open in T_1 then $p_1^{-1}(U)=U \times T_2$ is one of the open sets of $T_1 \times T_2$ in the basis named in Definition 3.5.1 (note that as usual we confuse T_2 with A_2). This proves that p_1 is continuous. The proof for p_2 is similar.

PROPOSITION 3.5.4. *Let* $f : T' \to T_1 \times T_2$ *be a map from a topological space* T' *to the topological product* $T_1 \times T_2$. *Then* f *is continuous iff* $p_1 \circ f$ *and* $p_2 \circ f$ *are continuous.*

Proof. If f is continuous then so are $p_1 \circ f$ and $p_2 \circ f$ by Propositions 3.5.3 and 3.1.8.

Suppose that $p_1 \circ f$ and $p_2 \circ f$ are continuous. For any subsets U_1,

U_2 of T_1, T_2 we have

$$\begin{aligned} f^{-1}(U_1\times U_2) &= f^{-1}((U_1\times T_2)\cap(T_1\times U_2)) \\ &= f^{-1}(U_1\times T_2)\cap f^{-1}(T_1\times U_2) \\ &= f^{-1}(p_1^{-1}(U_1))\cap f^{-1}(p_2^{-1}(U_2)) \\ &= (p_1\circ f)^{-1}(U_1)\cap(p_2\circ f)^{-1}(U_2). \end{aligned}$$

If U_1, U_2 are open in T_1, T_2 then $(p_1\circ f)^{-1}(U_1)$ and $(p_2\circ f)^{-1}(U_2)$ are open in T' by continuity of $p_1\circ f$ and $p_2\circ f$. Hence by the above, $f^{-1}(U_1\times U_2)$ is open in T'.

At this stage we have checked that $f^{-1}(W)$ is open for any set W in a certain basis for the topology on $T_1\times T_2$. It follows by Application 3.2.5 that f is continuous.

The discussion before Definition 3.5.1 showed further that the product topology is the only one which makes Proposition 3.5.4 true for all possible f.

The definition of topological products extends by induction to any finite number of factors. In particular it follows from Exercise 3.9.14 and induction that the product topology on $\mathbf{R}^n = \mathbf{R}\times\mathbf{R}\times\ldots\times\mathbf{R}$ (n factors) coincides with the topology arising from the Euclidean metric.

This is an appropriate point at which to mention how one would actually prove that $g:\mathbf{R}^2\to\mathbf{R}^2$ defined by $g(x, y)=(x^2+y^2, xy)$ is continuous (see Example 2.3.16(b)). By Proposition 3.5.4 it is enough to show that $h=p_1\circ g$ and $k=p_2\circ g$ are continuous. But for example $k(x, y)=xy$, so k is the product in the sense of Exercise 2.6.14 of the projections p_1 and p_2. Hence by that exercise k is continuous since p_1 and p_2 are continuous. Similarly h is continuous, so g is continuous.

The product topology may also be defined as the weak topology generated by the projections, or equivalently by naming as sub-basis all sets of the form $U_1\times T_2$ or $T_1\times U_2$ for U_1, U_2 open in T_1, T_2. This gives a smoother, but perhaps more sophisticated, development.

3.6 Homeomorphisms

This section touches on the more geometric aspects of topology.

DEFINITION 3.6.1. A *homeomorphism* $f: T_1\to T_2$ of topological spaces is a one–one correspondence such that f and f^{-1} are both continuous.

Equivalently, a homeomorphism f is a one–one correspondence such that U is open in T_1 iff $f(U)$ is open in T_2. In other words, f is a one–one correspondence which preserves all the structure that topological spaces possess (the analogous concept for algebraic structures such as groups or vector spaces is that of isomorphism). If there exists a homeomorphism $f: T_1 \rightarrow T_2$, we say that T_1 and T_2 are *homeomorphic* or *topologically equivalent*. It is an easy exercise to show that this is an equivalence relation.

EXAMPLES 3.6.2(a). Any two open intervals (a, b) and (c, d) in $\mathbf{R}$ are homeomorphic (where as usual $b>a$, $d>c$). The function $f:(a, b)\rightarrow(c, d)$ given by

$$f(x)=c+\{(d-c)(x-a)/(b-a)\}$$

is easily proved to be a homeomorphism, by writing down an explicit inverse function and noting that both f and f^{-1} are continuous.

(b) (a, b) and $\mathbf{R}$ are homeomorphic. By (a) it is enough to show that $(-1, 1)$ and $\mathbf{R}$ are homeomorphic. The function $f:(-1, 1)\rightarrow\mathbf{R}$ given by $f(x)=x/(1-|x|)$ is a homeomorphism.

(c) $[0, 1]$ and the subset of $\mathbf{R}^2$ consisting of $[0, 1]\times\{0\}\cup\{0\}\times[0, 1]$ are homeomorphic. We may refer to the latter subset as 'the letter L' provided we take L to be sanserif (i.e. L rather than L). We shall not construct an explicit homeomorphism, but suggest that it is obvious how one could be constructed.

(d) The letter L and the letter T (defined as in (c)) are not homeomorphic. Suppose that $f:\mathsf{T}\rightarrow\mathsf{L}$ were a homeomorphism, where the spaces themselves are their own names. Then for any x in T, the map g induced by f,

$$g: \mathsf{T}-\{x\}\rightarrow\mathsf{L}-\{f(x)\}$$

would also be a homeomorphism (see Exercise 3.9.17). If x is the junction point of T, then $\mathsf{T}-\{x\}$ consists of three disjoint half-open intervals. But no matter where $f(x)$ is in L, $\mathsf{L}-\{f(x)\}$ consists of one or two half-open intervals. The mathematics needed to prove the intuitively obvious fact that $\mathsf{T}-\{x\}$ and $\mathsf{L}-\{f(x)\}$ are not homeomorphic occurs in Chapter 6.

(e) A doughnut (solid ring) and a teacup (with handle) are homeomorphic. Again the existence of a homeomorphism is left to geometric intuition. This example gives rise to the definition of a topologist as a person who cannot tell the difference between these objects; they are indeed equivalent in topology in the same way that two congruent triangles are equivalent in Euclidean geometry.

(f) Consider the two closed curves in $\mathbf{R}^3$ shown in Fig. 3.1. The curve in Fig. 3.1(b) is called a trefoil knot (there are no breaks in it; the breaks in the figure are there to show which part of the curve goes behind the other). One minute with a piece of string will convince the reader that it *is* a knot, that is, it cannot be deformed continuously through $\mathbf{R}^3$ into a circle (this is quite hard to prove mathematically). Yet the two curves are homeomorphic. A homeomorphism may be defined by mapping the arc of the circle from x_i to x_{i+1} onto the piece of the trefoil knot from y_i to y_{i+1}, arrows agreeing.

There may well be many different homeomorphisms between two topological spaces. In Example 3.6.2(a), the formulae

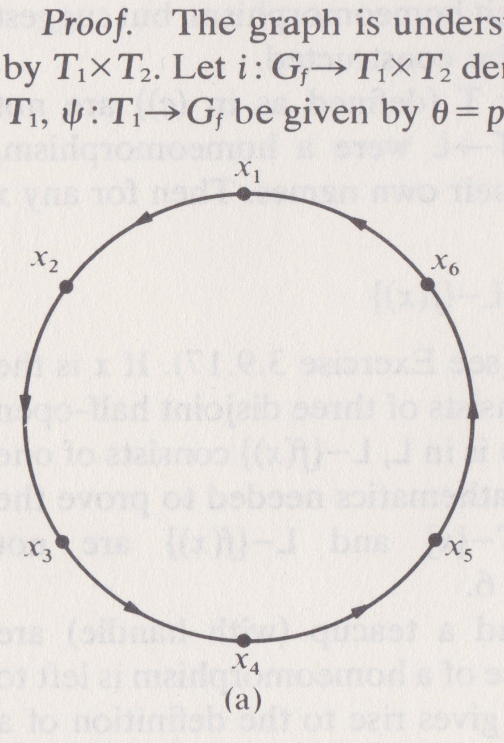
$$f_n(x) = c + \{(d-c)(x-a)^n/(b-a)^n\},$$

$$g_n(x) = d - \{(d-c)(x-a)^n/(b-a)^n\},$$

for any integer $n \geqslant 1$, give homeomorphisms from (a, b) to (c, d).

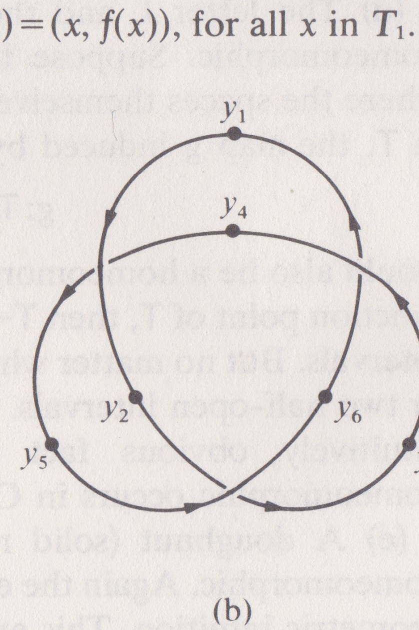
PROPOSITION 3.6.3. *If T_1, T_2 are topological spaces and $f: T_1 \to T_2$ is continuous, then the graph G_f of f is homeomorphic to T_1.*

Proof. The graph is understood to have the topology induced by $T_1 \times T_2$. Let $i: G_f \to T_1 \times T_2$ denote the inclusion map. Let $\theta: G_f \to T_1$, $\psi: T_1 \to G_f$ be given by $\theta = p_1 \circ i$, $\psi(x) = (x, f(x))$, for all x in T_1. It



(a)

FIG. 3.1(a)



(b)

FIG. 3.1(b)

is easily checked that θ and ψ are mutually inverse functions. Continuity of θ follows from continuity of p_1 and i. By Proposition 3.5.4, $i \circ \psi$ is continuous since $p_1 \circ i \circ \psi$ (the identity map of T_1), and $p_2 \circ i \circ \psi = f$ are continuous. Hence by Proposition 3.4.2(b), ψ is continuous.

If P is some property which makes sense for any topological space, and if whenever a topological space T has property P so has any topological space homeomorphic to T, we call P a topological property. If P mentions properties of individual points of T such as 'all the points in T are real numbers', then P will not be a topological property. For example $\mathbf{R}$ is clearly homeomorphic to the 'y-axis' in $\mathbf{R}^2$, that is, the subspace $\{(x, y) \in \mathbf{R}^2 : x = 0\}$, but the points in the latter space are not real numbers. But, roughly, apart from this kind of property, anything one can sensibly say about a topological space (without referring to extra structure such as a metric) is bound to be a topological property (to make this statement precise we should need to use mathematical logic). However, to be on the safe side, we shall formally check that various properties introduced in later chapters are topological properties.

A logically distinct concept is that of a topological property or invariant of spaces with extra structure, such as a metric. If P is some property which makes sense for any metric space, and if whenever a metric space has property P so has any homeomorphic metric space, we call P a topological invariant of metric spaces.

EXAMPLES 3.6.4. The property 'M contains exactly seven points' is a topological invariant (indeed, a set-theoretic invariant) since a homeomorphism is a one–one correspondence. We shall later see more exciting topological invariants than this.

The property 'M is bounded' makes sense for any metric space, but is not a topological invariant. For example the interval $(0, 1)$ is bounded while $\mathbf{R}$ is not, and these are homeomorphic.

3.7 Definitions

This section contains some more of the language of topological spaces. The definitions apply also to any metric space, via the topological space it gives rise to.

DEFINITION 3.7.1. A subset V of a topological space T is *closed in T* if $T - V$ is open in T.

EXAMPLES 3.7.2. Any 'closed interval' $[a, b]$ is closed in **R**. (Its complement is $(-\infty, a) \cup (b, \infty)$.) The interval $[a, b)$ is not closed in **R**, but it is closed in (c, b) for any $c < a$, and it is closed in **R** when **R** is given the discrete topology. Thus as usual we have to be precise about what space a set is closed in.

The names 'open' and 'closed' might suggest that any subset is one or the other and not both; both of these are false in general.

EXAMPLES 3.7.3. (a) The interval $(a, b]$ is neither open nor closed in **R**.

(b) Any topological space is both open and closed in itself.

Since the closed sets in a space are the complements of the open sets, the following dual properties are implied by the topological space axioms, using De Morgan's laws.

PROPOSITION 3.7.4. *In any topological space T, $\varnothing$ and T are closed, the union of two (and hence any finite number of) closed sets is closed, and the intersection of any family of closed sets is closed.*

As the reader may suspect, we could have stated the topological space axioms in terms of a fixed family of subsets called the closed sets and then defined an open set to be one whose complement is closed. There are further possibilities for axiom systems, but we do not pursue them.

One of these involves the useful concept in the next definition.

DEFINITION 3.7.5. A *neighbourhood* of a point x in a topological space T is a subset V of T such that $x \in U \subset V$ for some open subset U of T.

Some textbooks require neighbourhoods to be open, but our definition is more commonly used.

PROPOSITION 3.7.6. *Let T be a topological space and let $T' \subset T$ be a subspace. Then $V \subset T'$ is closed in T' iff $V = T' \cap W$ for some W closed in T.*

Proof. Suppose that $V \subset T'$ is closed in T'. Then $T' - V$ is open in T', so by definition of the subspace topology, $T' - V = T' \cap U$ for some U open in T. Then

$$V = T' - (T' - V) = T' - (T' \cap U) = T' \cap (T - U),$$

and $T - U$ is closed in T.

Conversely if $V = T' \cap W$ where W is closed in T, then $T' - V = T' - (T' \cap W) = T' \cap (T - W)$ is open in T', so V is closed in T'.

COROLLARY 3.7.7. *With the notation of Proposition 3.7.6, if T' is closed in T, then $V \subset T'$ is closed in T' iff V is closed in T.*

Proof. If $V \subset T'$ is closed in T then $V = T' \cap V$ is closed in T' by the proposition. If V is closed in T' then by the proposition $V = T' \cap W$ where W is closed in T. Since also T' is closed in T, V is closed in T by Proposition 3.7.4.

PROPOSITION 3.7.8. *A map $f: T_1 \rightarrow T_2$ of topological spaces is continuous iff for every V closed in T_2, $f^{-1}(V)$ is closed in T_1.*

Proof. This depends on the set-theoretic fact that for any $V \subset T_2$, $f^{-1}(T_2 - V) = T_1 - f^{-1}(V)$. Suppose that the condition on closed sets holds and that U is open in T_2. Then $T_2 - U$ is closed in T_2 and by hypothesis $f^{-1}(T_2 - U)$ is closed in T_1. Thus $T_1 - f^{-1}(U)$ is closed in T_1, so $f^{-1}(U)$ is open in T_1. Since this holds for any U open in T_2, f is continuous. The converse is proved similarly.

DEFINITION 3.7.9. Given a point x and a subset H of a topological space T, *x is a limit point of H* if every open set containing x contains some point of H other than x. In a metric space it is enough to insist that $B_\varepsilon(x) \cap H - \{x\} \neq \varnothing$ for all $\varepsilon > 0$, since any open set containing x contains $B_\varepsilon(x)$ for some $\varepsilon > 0$.

EXAMPLES 3.7.10. (a) In $\mathbf{R}$, a is a limit point of (a, b) and also of $[a, b]$. This illustrates that a limit point of a set may or may not be in the set.

(b) Let $H = \{0\} \cup (1, 2) \subset \mathbf{R}$. Then 0 is not a limit point of H (for example, $B_1(0)$ does not contain any point of H other than 0). The set of limit points of H in $\mathbf{R}$ is $[1, 2]$. This illustrates that a point in a set H may or may not be a limit point of H.

(c) Any point x in $\mathbf{R}$ is a limit point of $\mathbf{Q}$, since for any $\varepsilon > 0$, there exists a rational number in $(x, x + \varepsilon)$ for example. In this case $\mathbf{Q}$ is countable but its set of limit points in $\mathbf{R}$ is $\mathbf{R}$, which is uncountable.

DEFINITION 3.7.11. The *closure* $Cl(H)$ of a subset H in a topological space T is the union of H and all the limit points of H in T.

PROPOSITION 3.7.12. *A point x of T is in $Cl(H)$ iff every open set containing x contains a point of H.*

This follows from the definitions. In a metric space, it is equivalent to insist that $B_\varepsilon(x) \cap H \neq \varnothing$ for every $\varepsilon > 0$.

EXAMPLES 3.7.13. The closure of (a, b) or $(a, b]$ or $[a, b)$ or $[a, b]$ in **R** is $[a, b]$. The closure of $\{0\} \cup (1, 2)$ in **R** is $\{0\} \cup [1, 2]$. If $H \subset \mathbf{R}$ is non-empty and bounded above, then $u = \sup H \in Cl(H)$. For given $\varepsilon > 0$, there exists an x in $H \cap B_\varepsilon(u)$, by leastness of u. Similarly if H is non-empty and bounded below then $\inf H \in Cl(H)$.

EXAMPLES 3.7.14. (a) The closure of (a, b) in (a, ∞) is $(a, b]$.
(b) The closure of (a, b) in **R** with the discrete topology is (a, b).
Thus we have to be careful about what space closure is taken in.

In the next four propositions, everything takes place in some topological space T.

PROPOSITION 3.7.15.
(a) *H is closed in T iff $H = Cl(H)$,*
(b) $H \subset K \Rightarrow Cl(H) \subset Cl(K)$,
(c) $Cl(Cl(H)) = Cl(H)$,
(d) *$Cl(H)$ is closed.*

Proof. The definition of closure gives (b) immediately, and (d) follows from (a) and (c). To prove (a), suppose first that H is closed in T. Then for any x in $T - H$, $T - H$ is an open set containing x and not containing any point of H, so $x \notin Cl(H)$. This proves $Cl(H) \subset H$. Since $H \subset Cl(H)$ by definition of closure, $Cl(H) = H$. Conversely if $H = Cl(H)$ then for x in $T - H$, there is an open set U_x such that $x \in U_x$ and $U_x \cap H = \varnothing$, that is, $U_x \subset T - H$. Thus $T - H = \bigcup_{x \in T-H} U_x$ is open, so H is closed.

To prove (c) it is enough, by definition of closure, to prove $Cl(Cl(H)) \subset Cl(H)$. Let $x \in Cl(Cl(H))$. Then any open U containing x contains some point y in $Cl(H)$. Now viewing U as an open set containing y, we must have $U \cap H \neq \varnothing$. This proves that $x \in Cl(H)$, as required.

PROPOSITION 3.7.16. *Let K be the intersection of all closed sets containing H. Then $K = Cl(H)$. In other words, $Cl(H)$ is the smallest closed set containing H.*

Proof. If V is closed and $H \subset V$, then $Cl(H) \subset Cl(V) = V$. Hence $Cl(H) \subset K$. Conversely since $Cl(H)$ is closed and $H \subset Cl(H)$, it follows that $K \subset Cl(H)$. Hence $K = Cl(H)$.

PROPOSITION 3.7.17. *For any indexing set I,*

$$Cl \bigcap_I H_i \subset \bigcap_I Cl(H_i),$$

where $H_i \subset T$ for each i in I.

Proof. All intersections in the proof will be over *I*. Since $\bigcap Cl(H_i)$ is an intersection of closed sets, it is closed. Also, it contains $\bigcap H_i$, and therefore by Proposition 3.7.16 it contains $Cl(\bigcap H_i)$.

Equality does not necessarily hold in Proposition 3.7.17, even for a finite intersection (see Exercise 3.9.24).

PROPOSITION 3.7.18.

$$Cl(H_1 \cup H_2 \cup \ldots \cup H_n) = Cl(H_1) \cup Cl(H_2) \cup \ldots \cup Cl(H_n).$$

Proof. Let us write *K* for the right-hand side, and

$$H = H_1 \cup H_2 \cup \ldots \cup H_n.$$

For each $i = 1, 2, \ldots, n$, $H_i \subset H$, so $Cl(H_i) \subset Cl(H)$. Hence $K \subset Cl(H)$. This much of the argument works for any family of subsets $\{H_i\}$. Conversely *K*, being the union of a finite number of closed sets, is closed. Also, $K \supset H$. Hence by Proposition 3.7.16, $K \supset Cl(H)$ Hence $K = Cl(H)$.

As is to be expected, $Cl(H) \subset K$ is false in general if infinitely many sets are involved.

EXAMPLE 3.7.19. In **R** let $H_n = [1/n, 1]$ for $n \geqslant 2$. Then $Cl(H_n) = H_n$ and

$$\bigcup_{n \geqslant 2} Cl(H_n) = \bigcup_{n \geqslant 2} H_n = (0, 1],$$

while $Cl\left(\bigcup_{n \geqslant 2} H_n\right) = [0, 1]$.

DEFINITION 3.7.20. A subset *H* is (*everywhere*) *dense* in *T* if $Cl(H) = T$.

EXAMPLE 3.7.21. The rationals form a dense subset of **R** by Example 3.7.10(c).

It is a tempting fallacy to assume that in a metric space *M* with metric *d*, $Cl(B_\varepsilon(x)) = \{y \in M : d(x, y) \leqslant \varepsilon\}$. This is not necessarily true, although inclusion in one direction holds.

EXAMPLE 3.7.22. (a) In a discrete metric space M, $Cl(B_1(x))=\{x\}$, while $\{y\in M: d(x, y)\leq 1\}=M$.

(b) In the subspace $M=[0, 1]\cup[2, 3]$ of $\mathbf{R}$, $Cl(B_1(1))=[0, 1]$, while $\{y\in M: d(1, y)\leq 1\}=[0, 1]\cup\{2\}$.

PROPOSITION 3.7.23. *If $y\in Cl(B_\varepsilon(x))$ in a metric space with metric d, then $d(x, y)\leq\varepsilon$.*

Proof. Suppose that $d(x, y)>\varepsilon$. Then $B_{d(x,y)-\varepsilon}(y)$ is an open set containing y and not meeting $B_\varepsilon(x)$. Hence $y\notin Cl(B_\varepsilon(x))$.

The dual concept to closure is interior. We give the definition in the form dual to the conclusion of Proposition 3.7.16.

DEFINITION 3.7.24. The *interior* $\operatorname{Int} H$ of a subset H in a topological space T is the union of all subsets of H which are open in T.

EXAMPLES 3.7.25. $\operatorname{Int}[a, b]$ in $\mathbf{R}$ is (a, b). $\operatorname{Int}[a, b]$ in $(-\infty, b]$ is $(a, b]$. $\operatorname{Int}\mathbf{Q}$ in $\mathbf{R}$ is $\varnothing$.

DEFINITION 3.7.26. A subset H is *nowhere dense* in a topological space T if $\operatorname{Int}(Cl(H))=\varnothing$.

EXAMPLE 3.7.27. The set $\{1/n: n\in\mathbf{N}\}$ is nowhere dense in $\mathbf{R}$.

PROPOSITION 3.7.28. *A subset H of a topological space T is nowhere dense in T iff $T-Cl(H)$ is dense in T.*

Proof.

$$\begin{aligned}\operatorname{Int}(Cl(H))=\varnothing &\Leftrightarrow \text{every open subset of } T \text{ contains}\\ &\qquad \text{a point of } T-Cl(H)\\ &\Leftrightarrow T-Cl(H) \text{ is dense in } T.\end{aligned}$$

COROLLARY 3.7.29. *A closed subset H of a topological space T is nowhere dense in T iff $T-H$ is dense in T.*

The hypotheses that H is closed in T cannot be deleted here.

EXAMPLE 3.7.30. Let $T=\mathbf{R}$, $H=\mathbf{Q}$. Then $\mathbf{R}-\mathbf{Q}$ is dense in $\mathbf{R}$, but $\operatorname{Int}(Cl(\mathbf{Q}))=\mathbf{R}$.

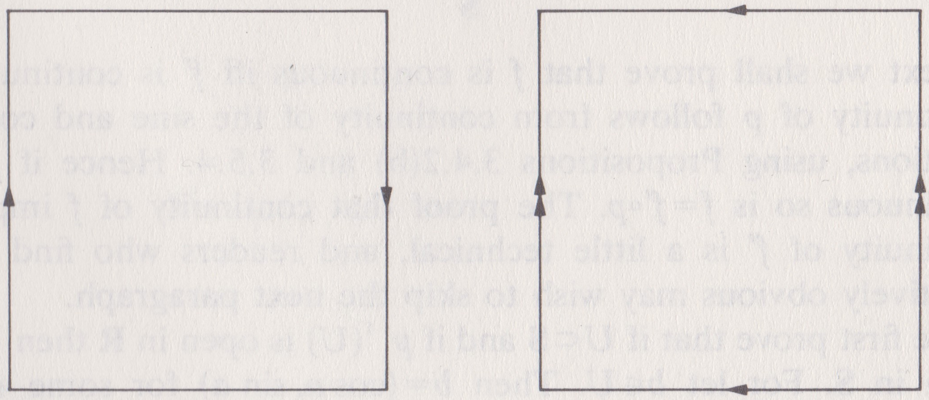
DEFINITION 3.7.31. The *boundary* (or *frontier*) $b(H)$ of a set H in a topological space T is $Cl(H)\cap Cl(T-H)$.

EXAMPLE 3.7.32. The boundary of $(0, 1]$ in $\mathbf{R}$ is $\{0, 1\}$. The boundary of $\mathbf{Q}$ in $\mathbf{R}$ is $\mathbf{R}$.

As the reader can imagine, many permutations of the definitions and propositions in this section are possible, and we have not catalogued all possible relations among the different concepts. It is not necessary to memorize them all before proceeding, but we shall refer back to them from time to time.

*3.8 Quotient spaces

It is quite likely that the reader has at some time taken a rectangular strip of paper, twisted one of the shorter sides through 180°, stuck it to the other short side, and gone to work on the resulting Möbius band with a pair of scissors for the entertainment of friends. The sticking part of this operation corresponds to forming an *identification* or *quotient* space T' from a space T. Any two (or in general more) points in T which get stuck together are said to be *identified* in T'. In set-theoretic terms (not yet worrying about topologies) there is a map p of T onto T' such that for any t' in T', all the points in $p^{-1}(t')$ get stuck together to form t'. Thus T is the disjoint union of all the sets $p^{-1}(t')$. To get the Möbius band as T', let T be the square $[0, 1]\times[0, 1]$. Then the decomposition of T into disjoint sets is rather simple. Each point $(0, y)$ goes together with the point $(1, 1-y)$ into a two-point set (since these are points which get stuck together in the Möbius band) while any point not on one of these vertical sides goes into a singleton set. The way in which the identification is made is usually indicated as in Fig. 3.2(a). A side marked with an arrow is stuck to any other side with the same kind of arrow, to make the arrows agree in direction. Likewise the torus (surface of a doughnut) is an identification



(a) (b)

FIG. 3.2(a)(b)

space of the square as shown in Fig. 3.2(b). We may think of this in two stages: first the single-arrowed sides are stuck together to form a circular cylinder and then the ends of this cylinder (which were formerly the double-arrowed sides of the square) are stuck together to form the torus.

We want to describe a topology on T' which in these examples should coincide with the topology induced on the Möbius band and the torus by the usual topology of $\mathbf{R}^3$. By looking carefully at the geometry here, one can guess how to topologize T' in general. We prefer to give a second example.

Recall that in Fourier series one is interested in functions $f:\mathbf{R}\to\mathbf{R}$ which are periodic, that is, for some fixed real number $2l$, $f(x+2l)=f(x)$ for every x in $\mathbf{R}$. For simplicity, let $l=\pi$. Such 2π-periodic functions may be produced as follows. Let $\mathbf{S}$ denote the circle of unit radius and centre the origin in $\mathbf{R}^2$. Let $p:\mathbf{R}\to\mathbf{S}$ be given by $p(x)=(\cos x, \sin x)$. Then for any function $f':\mathbf{S}\to\mathbf{R}$, the composition $f'\circ p$ is 2π-periodic, since, writing $f=f'\circ p$,

$$\begin{aligned} f(x+2\pi) &= f'(\cos(x+2\pi), \sin(x+2\pi)) \\ &= f'(\cos x, \sin x) = f(x). \end{aligned}$$

Moreover any 2π-periodic function $f:\mathbf{R}\to\mathbf{R}$ may be produced in this way: the formula $f'(\cos x, \sin x)=f(x)$ gives a well-defined function f' on $\mathbf{S}$ provided f is 2π-periodic. This sets up a one–one correspondence $f\leftrightarrow f'$, where the maps f, f' are related by $f=f'\circ p$. This relationship is expressed by saying that the following diagram *commutes:*

$$\begin{tikzcd} \mathbf{R} \arrow[rr, "f"] \arrow[dr, "p"'] & & \mathbf{R} \\ & \mathbf{S} \arrow[ur, "f'"'] & \end{tikzcd}$$

Next we shall prove that f is continuous iff f' is continuous. Continuity of p follows from continuity of the sine and cosine functions, using Propositions 3.4.2(b) and 3.5.4. Hence if f' is continuous so is $f=f'\circ p$. The proof that continuity of f implies continuity of f' is a little technical, and readers who find this intuitively obvious may wish to skip the next paragraph.

We first prove that if $U\subset\mathbf{S}$ and if $p^{-1}(U)$ is open in $\mathbf{R}$ then U is open in $\mathbf{S}$. For let $b\in U$. Then $b=(\cos a, \sin a)$ for some a in $p^{-1}(U)$. Since $p^{-1}(U)$ is open, $B_\delta(a)\subset p^{-1}(U)$ for some $\delta>0$. We may assume that $\delta<\pi$. Thus $b\in p(B_\delta(a))\subset U$. But $p(B_\delta(a))=\mathbf{S}\cap B_\varepsilon(b)$ as in Fig. 3.3, where an easy calculation shows that

$\varepsilon = 2\sin\frac{1}{2}\delta$. Thus $p(B_\delta(a))$ is an open ball in **S**, centred on b, and so U is open. Now suppose that f is continuous and let W be open in **R**. Then $p^{-1}(f'^{-1}(W)) = f^{-1}(W)$ is open in **R**, hence $f'^{-1}(W)$ is open in **S**, and so f' is continuous as required.

Thus the study of continuous 2π-periodic functions on **R** is reduced to the study of (ordinary) continuous functions on **S**.

More generally, let $T = \{A, \mathcal{T}\}$ be any topological space. We suppose given either a decomposition of A into disjoint subsets, or an equivalence relation on A. These amount to the same thing, since an equivalence relation on A divides A up into disjoint equivalence classes, and conversely any such decomposition of A gives rise to an equivalence relation $\sim$ on A, by letting $x \sim y$ iff x and y are in the same subset in the decomposition. Let A' denote the set of equivalence classes; thus there is just one point in A' for each set in the given decomposition of A. For any x in A, let $p(x)$ denote the point of A' which is the equivalence class in A to which x belongs. This defines a map $p: A \to A'$. Any map $f': A' \to B$ to a set B gives rise to a map $f = f' \circ p: A \to B$ such that $x \sim y \Rightarrow f(x) = f(y)$, i.e. f is constant on each equivalence class, and conversely any map $f: A \to B$ which is constant on each equivalence class is of the form $f' \circ p$ for some map $f': A' \to B$. We should like to topologize A' so that when B has a topology, f is continuous iff f' is continuous. It turns out that there is a unique topology on A' which fits the bill.

Suppose that such a topology $\mathcal{T}''$ on A' exists. First take $B = A'$ and $f = p$. Then f' is the identity map of A', which is continuous, so p is continuous. Thus if $U \in \mathcal{T}''$ then $p^{-1}(U) \in \mathcal{T}$, in other words $\mathcal{T}'' \subset \mathcal{T}'$, where $\mathcal{T}' = \{U \subset A' : p^{-1}(U) \in \mathcal{T}\}$.

Suppose that we can prove that $\mathcal{T}'$ is a topology for A'. Then take f' to be the identity map of A', thought of as a map from

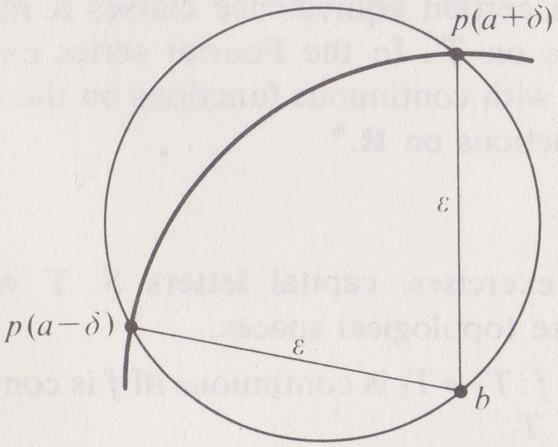


FIG. 3.3

$\{A', \mathcal{T}''\}$ to $\{A', \mathcal{T}'\}$. Since $f' \circ p = p$ is $(\mathcal{T}, \mathcal{T}')$-continuous, f' is $(\mathcal{T}'', \mathcal{T}')$-continuous. Hence $\mathcal{T}' \subset \mathcal{T}''$, so $\mathcal{T}'' = \mathcal{T}'$. The upshot is this: if we can prove that $\mathcal{T}'$ is a topology for A', then it is the unique topology fulfilling our original aim.

DEFINITION 3.8.1. Given a topological space $T = \{A, \mathcal{T}\}$, a set A', and a map p of A onto A', the *identification* or *quotient topology* on A' is

$$\mathcal{T}' = \{U \subset A' : p^{-1}(U) \in \mathcal{T}\}.$$

The check that $\mathcal{T}'$ is a topology is left as Exercise 3.9.39. The map p is called an *identification map.* It is continuous by Definition 3.8.1. The identification topology has been chosen to make the next proposition true.

PROPOSITION 3.8.2. *Given an identification map $p : T \to T'$, a topological space T'' and a commutative diagram of maps*

$$\begin{tikzcd} T \arrow[rr, "f"] \arrow[dr, "p"'] & & T'' \\ & T' \arrow[ur, "f'"'] & \end{tikzcd}$$

then f is continuous iff f' is continuous.

This result is dual to Proposition 3.4.2(b) in the sense that if all the arrows in the diagram are reversed, one gets the diagram for Proposition 3.4.2(b).

This correspondence $f \leftrightarrow f'$ proves fruitful in both directions. For example it is easier, at least notationally, to work with continuous maps with domain $[0, 1] \times [0, 1]$ and satisfying $f(0, y) = f(1, 1-y)$ for all y in $[0, 1]$, than to work with continuous maps with domain the Möbius band. Conversely, in place of maps on T which are constant on certain equivalence classes it may be easier to work with maps on T'. In the Fourier series example, we may choose to work with continuous functions on the circle instead of 2π-periodic functions on $\mathbf{R}$.*

Exercises 3.9

In all these exercises, capital letters S, T with or without subscripts denote topological spaces.

1. Prove that $f : T_1 \to T_2$ is continuous iff f is continuous at a for every a in T_1.

2. List all possible topologies on
   (i) a set $\{a, b\}$,
   (ii) a set $\{a, b, c\}$.
3. Give an example of topologies $\mathcal{T}_1$, $\mathcal{T}_2$ on a set A such that $\mathcal{T}_1 \not\subset \mathcal{T}_2$ and $\mathcal{T}_2 \not\subset \mathcal{T}_1$.
4. Prove that the intersection of two topologies on a set A is again a topology on A, but that their union may or may not be a topology on A. Does the intersection result extend to any family of topologies on A?
5. Prove Proposition 3.1.10.
6. Prove that any map $f: T_1 \rightarrow T_2$ is continuous if either T_1 is a discrete space or T_2 is an indiscrete space.
7. Prove that $f: T \rightarrow \mathbf{R}$ is continuous iff for every a in $\mathbf{R}$, $f^{-1}(-\infty, a)$ and $f^{-1}(a, \infty)$ are open in T.
8. Prove that the collection $\{B_q(s)\}$ of Example 3.2.7 is a countable basis for $\mathbf{R}^2$. Can you extend this result to $\mathbf{R}^n$?
9. Prove that Definition 3.4.1 gives a topology on H.
10. Prove that if H is open in T_1 and T_1 is open in T_2 then H is open in T_2.
11. Let $H \subset K \subset A$, let $\mathcal{T}$ be a topology for A, and let $\mathcal{T}_H$, $\mathcal{T}_K$ be the induced topologies on H, K. Prove that $\mathcal{T}_H$ coincides with the topology on H induced by $\mathcal{T}_K$.
12. Prove that if $\mathcal{B}$ is a basis for a topology $\mathcal{T}$ on a set A and if $H \subset A$, then $\mathcal{B}_H = \{H \cap B : B \in \mathcal{B}\}$ is a basis for the induced topology $\mathcal{T}_H$.
13. Given a metric space $\{A, d\}$ and a subset $H \subset A$, let $\mathcal{T}$ be the topology on A arising from d. Prove that the topology $\mathcal{T}_H$ induced by $\mathcal{T}$ on H coincides with the topology arising from the restricted metric d_H.
14. Suppose that the metric spaces $\{A, d\}$, $\{A', d'\}$ give rise to the topological spaces T, T'. Show that the topology on $A \times A'$ arising from any of the metrics d_1, d_2, d_∞ of Example 2.2.7 is the product topology of $T \times T'$.
15. Given maps $f_i : S_i \rightarrow T_i$ for $i = 1, 2$, let

$$f_1 \times f_2 : S_1 \times S_2 \rightarrow T_1 \times T_2$$

be defined by $(f_1\times f_2)(x, y)=(f_1(x), f_2(y))$. Prove that $f_1\times f_2$ is continuous iff f_1 and f_2 are continuous.

16. Prove that homeomorphism is an equivalence relation.
17. Prove that if $f: T_1\to T_2$ is a homeomorphism and $H\subset T_1$ then $g: H\to f(H)$ and $h: T_1-H\to T_2-f(H)$ induced by f are homeomorphisms.
18. Prove that the map $t: T_1\times T_2\to T_2\times T_1$ given by $t(x, y)=(y, x)$ is a homeomorphism.
19. Prove that if a, b are points in spaces T_1, T_2 then T_1, T_2 are homeomorphic to the subspaces $T_1\times\{b\}$, $\{a\}\times T_2$ of $T_1\times T_2$ respectively.
20. Prove that if f_1 and f_2 are homeomorphisms then so is $f_1\times f_2$.

21*. Show that $[0, 1]\times[0, 1)$ is homeomorphic to $[0, 1)\times[0, 1)$.

22. Prove that a finite subset of a metric space has no limit points.
23. Classify the capital letters of the alphabet up to homeomorphism, without rigorous proofs.
24. Give an example of open sets $H, K\subset \mathbf{R}$ such that $H\cap Cl(K)$, $Cl(H)\cap K$, $Cl(H)\cap Cl(K)$, and $Cl(H\cap K)$ are all different.
25. Prove that $f: T_1\to T_2$ is continuous iff $f(Cl(H))\subset Cl(f(H))$ for all $H\subset T_1$.
26. Prove that if $f: T_1\to T_2$ is continuous and $f(Cl(H))$ is closed then $f(Cl(H))=Cl(f(H))$.
27. Give a proof for or else a counterexample to each of the following assertions:

    (i) if $f: T_1\to T_2$ is continuous and $H\subset T_1$ then

    $$f(Cl(H))=Cl(f(H)),$$

    (ii) if $f: T_1\to T_2$ is continuous and $H\subset T_2$ then

    $$f^{-1}(Cl(H))=Cl(f^{-1}(H)),$$

    (iii) if H is open in T and $K\subset T$ then

    $$H\cap Cl(K)\subset Cl(H\cap K).$$

28. Suppose that $H\subset T_1\subset T_2$, where T_1 is a subspace of T_2, and let K_i denote the closure of H in T_i for $i=1, 2$. Prove that

(i) $K_1 = T_1 \cap K_2$,
(ii) if T_1 is closed in T_2 then $K_1 = K_2$.

29. Given $H \subset T$, prove that the boundary $b(H)$ is closed in T.

30. Given $H \subset T$, prove that
(i) H is closed in T iff $b(H) \subset H$,
(ii) $b(H) = \varnothing$ iff H is open and closed in T.

31. Describe the boundary of each of the following subsets in $\mathbf{R}$: $(0, 1)$, $\mathbf{Z}$, $\{1/n : n \in \mathbf{N}\}$, $\mathbf{Q} \cap [0, 1]$.

32. Prove that if H is a non-empty bounded subset of a metric space then $Cl(H)$ is bounded and diam $Cl(H) =$ diam H.

33. Show that $f: T_1 \to T_2$ is continuous if either
(i) $T_1 = \bigcup_{i=1}^{n} V_i$ where V_i is closed in T_1 and $f \mid V_i$ is continuous for $i = 1, 2, \ldots, n$, or
(ii) $T_1 = \bigcup_{I} U_i$ where I is any indexing set, U_i is open in T_1, and $f \mid U_i$ is continuous for each i in I.

34. (a) Prove that if W is open in $T_1 \times T_2$ then $p_1(W)$, $p_2(W)$ are open in T_1, T_2 respectively.
(b) Give an example of a closed set $W \subset \mathbf{R} \times \mathbf{R}$ such that $p_1(W)$ is not closed in $\mathbf{R}$.

35. Given $H \subset T_1$, $K \subset T_2$, prove that
(i) $\mathrm{Int}(H \times K) = \mathrm{Int}\, H \times \mathrm{Int}\, K$,
(ii) $Cl(H \times K) = Cl(H) \times Cl(K)$,
(iii) $b(H \times K) = (b(H) \times Cl(K)) \cup (Cl(H) \times b(K))$.

36. Prove that in the space $\mathscr{C}[0, 1]$ of Example 2.2.8, $Cl(B_1(0)) = \{f \in \mathscr{C}[0, 1] : d_\infty(f, 0) \leq 1\}$, where 0 denotes the constant function with value zero.

37. Prove that any metric space which contains a countable dense set is second countable. Is the converse true?

38. By considering intervals in $\mathbf{R}$, show that T_1 may be homeomorphic to a subspace of T_2 and T_2 homeomorphic to a subspace of T_1 without T_1 and T_2 being homeomorphic.

39. Prove that Definition 3.8.1 gives a topology for A'.

40. Prove that for any topological product $T_1 \times T_2$, the projection maps p_1 and p_2 are identification maps.

41. Show that the composition of two identification maps is an identification map.

4

The Hausdorff condition

4.1 Motivation

For many, though not all, applications of topological spaces it is appropriate to impose a certain condition in addition to the topological space axioms. To motivate this condition, let us try to generalize the idea of a convergent sequence of real numbers. We shall go via metric spaces to make the definition plausible. As in § 1.2, a sequence of objects in any set A may be defined formally as a map $s: \mathbf{N} \to A$, but we use the traditional notation (x_n) for a sequence, putting $s(n) = x_n$.

Definition 4.1.1. A sequence (x_n) of points in a metric space M with metric d *converges* to a point x of M if given any (real number) $\varepsilon > 0$, there exists (an integer) N such that $x_n \in B_\varepsilon(x)$ for all $n \geqslant N$.

As we know from Chapters 2 and 3, the way to generalize this to topological spaces is to replace open balls by open sets.

Definition 4.1.2. A sequence (x_n) of points in a topological space T *converges* to a point x in T if given any open set U containing x, there exists N such that $x_n \in U$ for all $n \geqslant N$.

However, convergence in a general topological space does not necessarily behave in a way that can be reconciled with our intuition.

Example 4.1.3. Let T be an indiscrete space (Example 3.1.6). Then any sequence (x_n) in T converges to any point x of T. For given any open set U containing x, we must have $U = T$, so $x_n \in U$ for all $n \geqslant 1$.

Let us try to pinpoint what has led to this, by considering how one proves the uniqueness of limits of real number sequences. If we analyze the proof of Proposition 1.2.3, we see that it may be stated as follows. Suppose that (x_n) converges to x and also to x' (in

R). If $x \neq x'$, then we may choose $\varepsilon > 0$ (say $\varepsilon = \frac{1}{2}|x - x'|$) so that $B_\varepsilon(x)$ and $B_\varepsilon(x')$ are disjoint, and since x_n is supposed to belong to both of these neighbourhoods for all sufficiently large n, we get a contradiction. In Example 4.1.3, on the other hand, there are no disjoint open sets U, U' containing x, x'.

4.2 Separation axioms

DEFINITION 4.2.1. A topological space T *satisfies the Hausdorff condition* if given any two distinct points x, y in T there exist disjoint open subsets U, V of T containing x, y respectively. We shall refer to a topological space satisfying the Hausdorff condition as a *Hausdorff space.*

A distinguished professor used to embed the definition in his students' minds by saying that a space is Hausdorff if any two distinct points can be housed off from each other by disjoint open sets.

PROPOSITION 4.2.2. *In a Hausdorff space, any given convergent sequence has a unique limit.*

For the proof, one simply replaces the open balls $B_\varepsilon(x)$ and $B_\varepsilon(x')$ in § 4.1 by disjoint open sets containing x, x' respectively.

EXAMPLE 4.2.3. Any metric space $\{A, d\}$ is Hausdorff. For if $x, y \in A$ with $x \neq y$, then $d(x, y) > 0$. Put $\varepsilon = \frac{1}{2}d(x, y)$. Then $B_\varepsilon(x)$ and $B_\varepsilon(y)$ are disjoint open sets containing x, y respectively (see Exercise 2.6.1).

The proofs of the following properties are left as Exercise 4.3.2.

PROPOSITION 4.2.4. (a) *Any subspace of a Hausdorff space is Hausdorff.*

(b) *The topological product of two Hausdorff spaces is Hausdorff.*

(c) *If $f: T_1 \to T_2$ is injective and continuous, and if T_2 is Hausdorff then so is T_1.*

(d) *Hausdorffness is a topological property.*

We mention without proof that a quotient space of a Hausdorff space may not be Hausdorff (see, for example, p. 98 of Kelley (1955)).

The Hausdorff condition is just one of a hierarchy of distinct separation axioms or Tychonoff conditions which a topological space may or may not satisfy. There is some disagreement about the names of the various conditions, including the following two samples.

DEFINITIONS 4.2.5. A topological space is *regular* (*normal*) if given any closed subset V and x in $T-V$ (closed subset $V'\subset T-V$), there exist disjoint open subsets U, U' of T such that $V\subset U$, $x\in U'$ ($V'\subset U'$).

Exercises 4.3

1. Show that the Zariski topology of Example 3.1.7 is not Hausdorff.
2. Prove Proposition 4.2.4.
3. Prove that if $f: T_1\to T_2$ is a continuous map of a topological space T_1 to a Hausdorff space T_2, then the graph G_f is a closed subset of $T_1\times T_2$.
4. (a) Prove that if x is a point in a Hausdorff space T, then the intersection of all open subsets of T containing x is $\{x\}$.
   (b) Give an example to show that the conclusion of (a) does not imply that T is Hausdorff. (Hint: Exercise 1.)

5*. Let M be a metric space with metric d, and for any non-empty subset K and point x in M, let $d(x, K)=\inf\limits_{y\in K} d(x, y)$. Prove that
   (a) if K is closed in M then $d(x, K)=0$ iff $x\in K$,
   (b) $d(x, K)\leqslant d(x, y)+d(y, K)$ for any x, y in M,
   (c) the map $f_K: M\to \mathbf{R}$ given by $f_K(x)=d(x, K)$ is continuous.
   Suppose that H, K are disjoint closed sets in M. Let $g: M\to\mathbf{R}$ be given by $g=f_K-f_H$. Prove that $g^{-1}(0,\infty)$ and $g^{-1}(-\infty, 0)$ are disjoint open sets containing H, K respectively. (This shows that any metric space is normal.)

6. Let $f, g: T_1\to T_2$ be continuous maps of a topological space T_1 to a Hausdorff space T_2. Show that $W=\{x\in T_1: f(x)=g(x)\}$ is closed in T_1. Deduce that if $f: T\to T$ is a continuous map of a Hausdorff space T, then the 'fixed-point set' $\{x\in T: f(x)=x\}$ is closed in T.

5

Compact spaces

5.1 Motivation

THE subject matter of this chapter is probably the most important single topic in this book. There are several ways of framing the definition of compactness. The definition in this chapter is appropriate for topological spaces. Another important definition which works well in metric spaces will be studied in Chapter 7 and related to the present definition.

Recall from the introduction that we are aiming to prove some basic results about continuous functions in a general setting, and that the following is an example of the kind of result we wish to generalize.

PROPOSITION 5.1.1. *A continuous function* $f:[a, b]\to\mathbf{R}$ *is bounded on* $[a, b]$.

We shall begin with a slow build-up towards one way of proving this, and generalizations of it. Let us set out by supposing that the function $f:A\to\mathbf{R}$ is defined on some general subset A of $\mathbf{R}$. We may ask, 'Is f bounded on A?', that is, 'Does there exist a fixed real number K such that $|f(x)|\leq K$ for all x in A?' We shall proceed from the known to the unknown in examining this question.

STEP 1. If A is a finite set, say $A=\{a_1, a_2, \ldots, a_r\}$, then the answer is 'Yes', and we may take

$$K=\mathrm{Max}\{|f(a_1)|, |f(a_2)|, \ldots, |f(a_r)|\}.$$

STEP 2. More generally, if A is a finite union of subsets, say $A=\bigcup_{i=1}^{r} A_i$, and if we already know that f is bounded on each A_i separately, say $|f(x)|\leq K_i$ for all x in A_i, then again f is bounded on A: we set $K=\mathrm{Max}\{K_1, K_2, \ldots, K_r\}$, and then for any x in A, $x\in A_i$ for some i, so $|f(x)|\leq K_i\leq K$.

EXAMPLE 5.1.2. We now look at an example of a function which is not bounded. Let $A=(0, 1)$ and put $f(x)=1/x$ for all x in $(0, 1)$. Given any real number $K>0$, we can find a real number x such that $0<x<1/K$, (and $x<1$) and then $f(x)=1/x>K$. So no K is large enough to bound $f(x)$ on all of $(0, 1)$. This illustrates that the answer to our question can be negative even when f is continuous.

STEP 3. However, we get *something* when f is continuous. For suppose that $f: A\rightarrow \mathbf{R}$ is continuous, and for any a in A let us apply the $\varepsilon-\delta$ definition of continuity with $\varepsilon=1$, say. Thus there exists some $\delta>0$ such that $|f(x)-f(a)|<1$ for all x in A satisfying $|x-a|<\delta$. So $|f(x)|<1+|f(a)|$ for all x in $B_\delta(a)$. The δ here in general depends on a and on f as well as on the chosen value 1 of ε. Let us write it in the meantime as $\delta(a)$. So for continuous f, given any a in A, there is a single bound $K_a=1+|f(a)|$ for $|f(x)|$ which serves for all x in some neighbourhood $B_{\delta(a)}(a)$ of a.

STEP 4. Now recall that the original question is whether there is a single bound for $|f(x)|$ which serves for all x in A. We cannot in general answer this affirmatively by taking the maximum of the K_a in Step 3, because in general there are infinitely many K_a involved (one for each a in A) and the set of these K_a may not be bounded, far less have a maximum. However, suppose for a moment that A is the union of some *finite* number of the $B_{\delta(a)}(a)$ occurring in Step 3. Then, since f is bounded on each $B_{\delta(a)}(a)$, it follows by Step 2 that f is bounded on A.

Let us examine more closely the assumption which allowed us to reach this conclusion. Originally we just had $A=\bigcup_{a\in A} B_{\delta(a)}(a)$, and we then assumed that finitely many of these neighbourhoods were sufficient to give A as their union. Now the neighbourhoods $B_{\delta(a)}(a)$ depended on f as well as on a, so if we want to set down a condition *on* A which will enable us to prove by the above argument that *any* continuous real-valued function on A is bounded, we had better assume something like the property defined as follows.

PROVISIONAL DEFINITION 5.1.3. A subset A of $\mathbf{R}$ is *compact* if whenever A is the union of a collection of balls open in A, it is a union of finitely many of the balls in that collection.

Before stating the definition more precisely, let us consider what compactness enables us to do. The conclusion of Step 3, that any continuous real-valued function is bounded on some neighbourhood of each point in its domain, may be called a local statement, because it makes an assertion only about some neighbourhood of each point. The statement that f is bounded on A, however, may be called a global statement because it describes a property of f on the whole domain A. Compactness of A allows us to pass from the local to the global.

As it stands, Definition 5.1.3 makes sense for any metric space. As usual, replacing open balls by open sets generalizes it to topological spaces.

5.2 Definition of compactness

The definition of compactness may conveniently be expressed in the language of covers.

DEFINITIONS 5.2.1. A *cover* for a set A is a collection $\mathcal{U}$ of sets such that $A \subset \bigcup_{U \in \mathcal{U}} U$. A *subcover* of a given cover $\mathcal{U}$ for A is a subcollection $\mathcal{V} \subset \mathcal{U}$ which still forms a cover for A. A cover $\mathcal{U}$ is *finite* if it is finite, that is, if there are only finitely many sets in the collection $\mathcal{U}$. If A is a subspace of a topological space T, and if $\mathcal{U}$ is a cover for A in which each set U is open in T, then we say that $\mathcal{U}$ is an *open* cover for A (in particular A can equal T in this definition).

DEFINITION 5.2.2. A topological space T is *compact* if every open cover of T has a finite subcover.

It is important to notice exactly what this definition says: given any open cover $\mathcal{U}$ of T, there exists a finite number of the open sets in $\mathcal{U}$ which are enough to cover T. This is very different, for example, from saying that there exists a finite open cover of T. Indeed, the latter assertion is trivially true for any topological space T, for we may take the singleton collection $\{T\}$ as a finite open cover.

To emphasize the definition further, let us see that $(0, 1)$ is not compact. Consider the collection of open intervals $(1/n, 1)$ for all integers $n \geqslant 2$. Each of these is open in $(0, 1)$, and $(0, 1) = \bigcup_{n=2}^{\infty} (1/n, 1)$. (Note that given any x in $(0, 1)$, there exists an integer n such that $n > 1/x$, and then $x \in (1/n, 1)$.) But no finite subcollection of this particular collection will suffice to cover $(0, 1)$, since the union of any finite subcollection

$$\{(1/n_1, 1), (1/n_2, 1), \ldots, (1/n_r, 1)\}$$

is just $(1/N, 1)$, where $N = \mathrm{Max}\{n_1, n_2, \ldots, n_r\}$.

On the other hand, we shall shortly see that a closed bounded interval $[a, b]$ in $\mathbf{R}$ is compact. As an example for the meantime, it is easy to see that any topological space containing only finitely many points is compact.

Now that we have a definition of compactness, we may ask 'What is it good for ?' There are at least two general answers which are worth thinking about.

(1) It allows us to pass from the local to the global in the sense illustrated in § 5.1. In particular, the discussion in § 5.1 amounted to a proof that any real-valued continuous function on a compact space is bounded. (The proof will be repeated shortly.)

(2) The second answer has been very well expressed in Hewitt (1960). Hewitt remarks that compactness is a substitute for finiteness, appropriate for the analysis of continuity. More explicitly, he points out that many statements about functions $f: A \to B$ are:

  (i) true and trivial if A is a finite set,
  (ii) true for continuous f when A is a compact space,
  (iii) false, or very hard to prove, even for continuous f, when A is non-compact.

We have already seen this illustrated for the statement 'f is bounded on A', in §5.1, with $B = \mathbf{R}$. Later we shall see that another such statement is 'f attains its bounds on A'.

†For readers who know about vector spaces, it is worth mentioning that, in a similar sense, finite-dimensionality of a vector space is a substitute for finiteness, appropriate for the study of linear transformations. Consider for example the proposition that a linear transformation from a finite-dimensional vector space to itself is injective iff it is onto.†

REMARKS 5.2.3. (a) Suppose that T is a subspace of another topological space S. Definition 5.2.1 says that T is compact if any cover of T by subsets of T open in T has a finite subcover. Figure 5.1(a) shows what such a subcover might look like in a special case. We shall see that the following is an equivalent definition.

DEFINITION 5.2.4. With the above notation, T is *compact* if given any cover $\mathcal{U}$ of T by sets open in S, there is a finite subcover of $\mathcal{U}$ for T.

Figure 5.1(b) shows what such a finite subcover might look like in the same special case as Fig. 5.1(a).

Since it is often Definition 5.2.4 which we use in practice, we shall prove the equivalence of the definitions in detail. Suppose that T is compact according to Definition 5.2.2, and let $\mathcal{U}$ be a cover of T by subsets of S open in S. Then for each U in $\mathcal{U}$, $U \cap T$ is open in T, by definition of the subspace topology. Since $\mathcal{U}$ covers T, it is easy to see that $\{U \cap T : U \in \mathcal{U}\}$ covers T. Hence by hypothesis there exists a finite subcover, say $\{U_1 \cap T, U_2 \cap T, \ldots, U_r \cap T\}$, and so $\{U_1, U_2, \ldots, U_r\}$ is a finite subcover of $\mathcal{U}$ for T. Hence T is compact according to Definition 5.2.4.

Conversely, suppose that T is compact according to Definition 5.2.4 and let $\mathcal{V}$ be a cover for T by subsets of T open in T. Then by definition of the subspace topology, each V in $\mathcal{V}$ is of the form $U \cap T$ for some U open in S. Clearly the collection of all such U is a cover of T by sets open in S, so by hypothesis there is a finite subcover, say $\{U_1, U_2, \ldots, U_r\}$. Then $\{V_1, V_2, \ldots, V_r\}$, where $V_i = U_i \cap T$, is a finite subcover of $\mathcal{V}$ for T, and T is compact in the sense of Definition 5.2.2.

(b) In trying to see how a proof concerning compactness should go, it is sometimes helpful to experiment with pictures in $\mathbf{R}^2$, although of course such a picture does not give a rigorous proof.

(c) The mathematician H. Weyl is credited with a striking comment about compact subsets of the plane: 'If a city is compact, it can be guarded by a finite number of arbitrarily near-sighted policemen'.

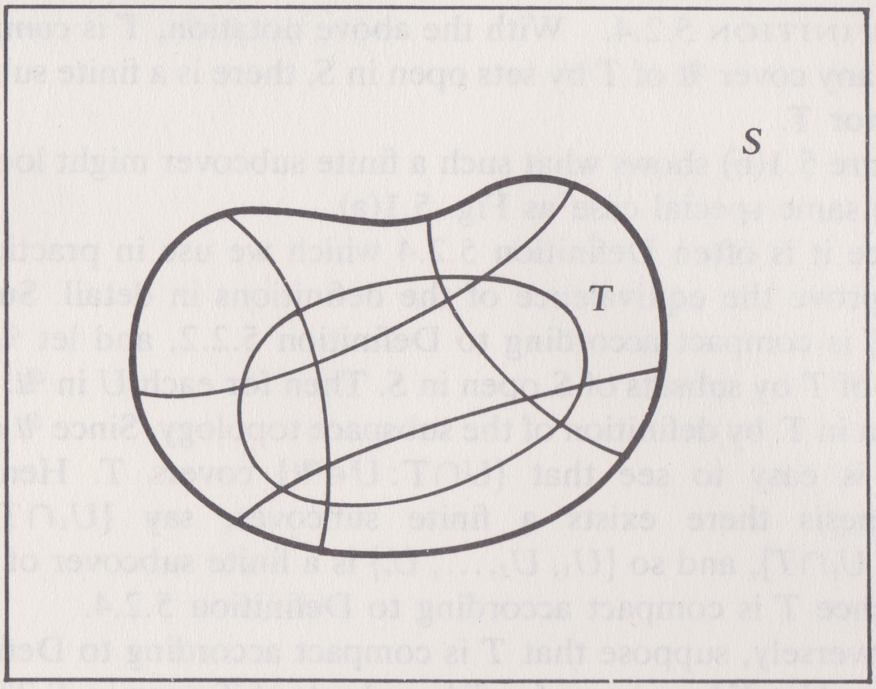
(d) In some textbooks, compactness is taken to mean the condition expressed in Definition 5.2.2 plus the Hausdorff condition.

(e) We abbreviate 'compact topological space' to 'compact space'.

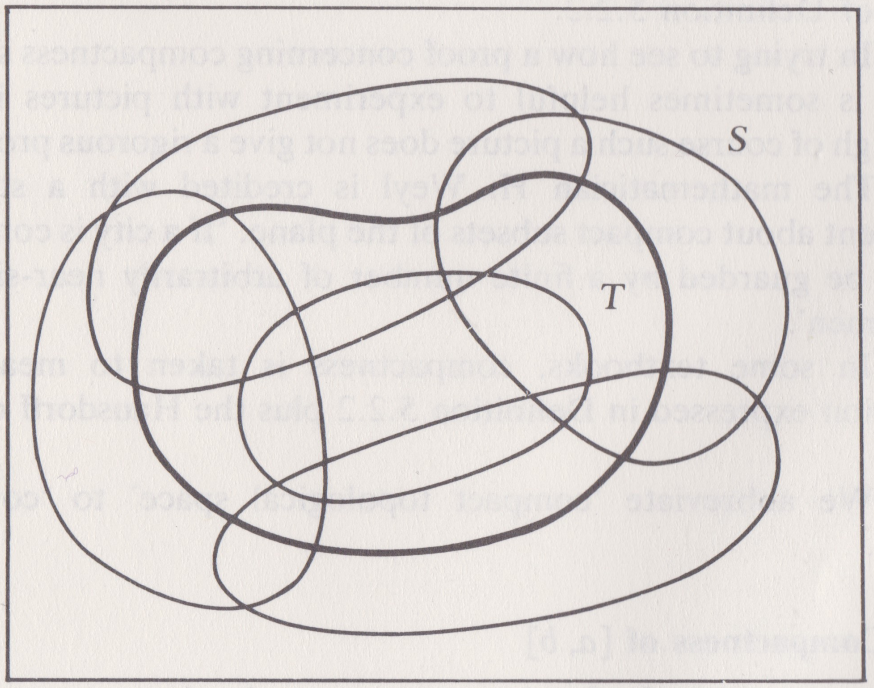
5.3 Compactness of $[a, b]$

We now come to the first important example of a compact space.

THEOREM 5.3.1 (Heine–Borel). *Any closed bounded interval* $[a, b]$ *in* $\mathbf{R}$ *is compact.*



(a)



(b)

FIG. 5.1(a)(b)

There are various ways of proving this. The proof below has been chosen for the direct way in which it uses the completeness axiom. Another popular proof, which may be preferred by some, is suggested in Exercise 5.10.16, and Chapter 10 contains yet another variant of the proof. It is valuable to work through different proofs.

Proof. Suppose that $\mathcal{U}$ is an arbitrary open cover of $[a, b]$ by sets open in $\mathbf{R}$ (cf. Remark 5.2.3(a)). Let $G=\{x\in\mathbf{R}: x\geqslant a$, and $[a, x]$ is covered by a finite subcollection of $\mathcal{U}\}$. Let us call points in G *good (for* $\mathcal{U}$). We want to show that b is good. Note that if x is good and if $a\leqslant y\leqslant x$ then y is good: for then $[a, y]\subset[a, x]$, so if $[a, x]$ is covered by a finite subcollection of the sets in $\mathcal{U}$, then $[a, y]$ is covered by the same finite subcollection.

We next show that $G\neq\varnothing$. It is convenient to show at the same time that $G\supset[a, a+\delta)$ for some $\delta>0$. To prove this, notice that a must belong to some U in $\mathcal{U}$, since $\mathcal{U}$ covers $[a, b]$. Since U is open, $[a, a+\delta)\subset U$ for some $\delta>0$. Hence $[a, x]\subset U$ for all x in $[a, a+\delta)$, and since this means that $[a, x]$ is covered by *one* set from $\mathcal{U}$, all such x are good.

Now the non-empty set G is either bounded above or not. If it is not, then there is some good c with $c>b$, and by our initial observation, it follows that b also is good.

We may therefore suppose that G is bounded above. Then by the completeness axiom (Axiom 1.1.4) we may let $g=\sup G$. If $g>b$, then there is some good c with $c>b$, by leastness of g, and again by our initial observation b is good. Suppose, to get a contradiction, that $g\leqslant b$. Notice that $g>a$, since we have shown that $[a, a+\delta)\subset G$ for some $\delta>0$. Since $g\in[a, b]$, g must belong to some U_0 in $\mathcal{U}$. Since U_0 is open, $U_0\supset B_\varepsilon(g)$ for some $\varepsilon>0$ (and since $g>a$, we may suppose also that $\varepsilon<g-a$). Now by leastness of g, there is a good c with $c>g-\varepsilon$. This means that $[a, c]$ is covered by a finite subcollection of $\mathcal{U}$, say $\{U_1, U_2,\ldots, U_r\}$. Then $[a, g+\frac{1}{2}\varepsilon]$ is covered by $\{U_1, U_2,\ldots, U_r, U_0\}$. Hence $g+\frac{1}{2}\varepsilon$ is good, contradicting the fact that g is an upper bound for G. This contradiction implies that $g>b$, and the proof is complete.

It is instructive to examine where this proof breaks down if one tries to apply it to $(0, 1]$ or $[0, 1)$.

Taking Theorem 5.3.1 along with § 5.1, we now have a proof of Proposition 5.1.1. The route we have followed is not the fastest one from the completeness axiom to a proof of this result. The aim here has been to shed some light on the definition of compactness.

5.4 Properties of compact spaces

We now study further the question of which spaces are compact. In particular, which subspaces of $\mathbf{R}^n$ are compact? In this section we show that for a subspace of a metric space to be compact, it is necessary that it be closed and bounded. In § 5.7 we shall show that for subspaces of $\mathbf{R}^n$ these conditions are also sufficient. We shall see, however, that they are not sufficient in more general metric spaces.

PROPOSITION 5.4.1. *Any compact subspace C of a metric space M is bounded.*

Proof. Let a be any point of M. Then $C \subset \bigcup_{n=1}^{\infty} B_n(a)$, since for any x in C, $d(x, a) < n$ for some integer n. Thus the collection $\{B_n(a) : n \in \mathbf{N}\}$ forms an open cover for C, and since C is compact, there is a finite subcover, say $\{B_{n_1}(a), B_{n_2}(a), \ldots, B_{n_r}(a)\}$. Let $n = \mathrm{Max}\{n_1, n_2, \ldots, n_r\}$. Then $C \subset \bigcup_{i=1}^{r} B_{n_i}(a) = B_n(a)$, and the proof is complete.

The integers in this proof could be replaced by any unbounded set of positive numbers.

PROPOSITION 5.4.2. *Any compact subspace C of a Hausdorff space T is closed in T.*

Proof. Let $a \in T - C$. We shall prove that there exists an open set U_a such that $a \in U_a \subset T - C$. Then $T - C = \bigcup_{a \in T-C} U_a$, so $T - C$ is open and hence C is closed.

Now for any single point x of C, the Hausdorff condition ensures the existence of disjoint open sets, say $U(x)$, $V(x)$, containing a, x respectively. If there were only a finite number of points $x_1, x_2, \ldots, x_r$ in C, we could take $U_a = \bigcap_{i=1}^{r} U(x_i)$, and get $a \in U_a \subset T - C$. (A diagram drawn by the reader will help in seeing this.) However, even if C is not finite, compactness of C is good enough. For the collection $\{V(x) : x \in C\}$ is an open cover of C, and hence has a finite subcover, say $\{V(x_1), V(x_2), \ldots, V(x_r)\}$. Let $U_a = \bigcap_{i=1}^{r} U(x_i)$. Then U_a, as a finite intersection of open sets, is open. Also, $a \in U_a$ since

$a \in U(x_i)$ for every $i = 1, 2, \ldots, r$. Finally, if $b \in U_a$ then for any $i = 1, 2, \ldots, r$, $b \in U(x_i)$ and hence $b \notin V(x_i)$, so $b \notin C$, since $C \subset \bigcup_{i=1}^{r} V(x_i)$. Thus $U_a \subset T - C$, as required.

It follows from Proposition 5.4.2 that singleton point sets (and more generally finite subsets) in a Hausdorff space are closed.

DEFINITION 5.4.3. A subspace T of a topological space S is *relatively compact* in S if $Cl(T)$ is compact.

For example (0, 1) is relatively compact in **R**, but not in (0, 1). Care should be taken not to confuse relative compactness with the alternative Definition 5.2.4 of compactness.

5.5 Continuous maps on compact spaces

PROPOSITION 5.5.1. *If $f: T_1 \to T_2$ is a continuous map of topological spaces and T_1 is compact then $f(T_1)$ is compact.*

Proof. Suppose that $\mathcal{U}$ is an open cover of $f(T_1)$, by sets open in T_2. By continuity of f, $f^{-1}(U)$ is open in T_1 for every U in $\mathcal{U}$. The collection $\{f^{-1}(U) : U \in \mathcal{U}\}$ is an open cover of T_1, since for any x in T_1, $f(x)$ must belong to some U in $\mathcal{U}$. By compactness of T_1, there is a finite subcover $\{f^{-1}(U_1), f^{-1}(U_2), \ldots, f^{-1}(U_r)\}$, and it is easily checked that $\{U_1, U_2, \ldots, U_r\}$ is then a finite subcover of $\mathcal{U}$ for $f(T_1)$.

This proposition is often stated as: a continuous image of a compact space is compact.

COROLLARY 5.5.2. *Compactness is a topological property.* (Recall what this means: if T_1 and T_2 are homeomorphic then T_1 is compact iff T_2 is compact.)

COROLLARY 5.5.3. *Any continuous map from a compact space to a metric space is bounded.*

Proof. This follows from Propositions 5.5.1 and 5.4.1. It reproves Proposition 5.1.1 in a general setting.

If a map $f: C \to \mathbf{R}$ is bounded, this means that $f(C)$ is a bounded subset of **R**, so $\sup f(C)$ and $\inf f(C)$ exist. These are called the bounds of f on C. In general they need not themselves be in $f(C)$, but if they are, then we say that f attains its bounds on C.

COROLLARY 5.5.4. *If $f: C \to \mathbf{R}$ is continuous and C is a compact space, then f attains its bounds on C.*

Proof. By Corollary 5.5.3, $f(C)$ is bounded. By Example 3.7.13, $\sup f(C)$ and $\inf f(C)$ belong to $Cl(f(C))$. By Propositions 5.5.1 and 5.4.2, $f(C)$ is closed in $\mathbf{R}$, hence $\sup f(C)$ and $\inf f(C)$ belong to $f(C)$, in other words f attains its bounds on C.

In particular, any continuous real-valued function on a closed bounded interval $[a, b]$ attains its bounds.

We shall study two more properties of continuous functions on compact spaces in §§ 5.8, 5.9.

5.6 Compactness and constructions

Before proving in § 5.7 that closed bounded subsets of $\mathbf{R}^n$ are compact, we need to study compactness for subspaces and products.

PROPOSITION 5.6.1. *Any closed subspace C of a compact space T is compact.*

Proof. Let $\mathcal{U}$ be any cover of C by sets open in T. Since C is closed, $T-C$ is open, and if we add $T-C$ to the collection $\mathcal{U}$ we get an open cover of T. By compactness of T, there is a finite subcover, say $\{U_1, U_2, \ldots, U_r\}$. This *a fortiori* covers C, and moreover if $T-C$ is one of these U_i, then we may throw it out, and the remaining $r-1$ sets will still cover C. This provides a finite subcover of $\mathcal{U}$ for C, as required.

THEOREM 5.6.2. *A topological product $T_1 \times T_2$ is compact iff T_1 and T_2 are both compact.*

Proof. In one direction, the implication is easy. If $T_1 \times T_2$ is compact then since T_1 and T_2 are the continuous images of $T_1 \times T_2$ by the projections p_1 and p_2, they are compact by Proposition 5.5.1.

Now suppose that T_1 and T_2 are compact. The proof that $T_1 \times T_2$ is compact is one of the trickier proofs in topology at this level. We shall take it in easy stages, in the manner of § 5.1.

Let $\mathcal{W}$ be an arbitrary open cover of $T_1 \times T_2$. We shall call a subset $A \subset T_1$ *good (for $\mathcal{W}$)* if $A \times T_2$ is covered by a finite subcollection of $\mathcal{W}$. We want to prove that T_1 is good.

STEP 1. If $A_1, A_2, \ldots, A_r \subset T_1$ are all good, then so is $A = A_1 \cup A_2 \cup \ldots \cup A_r$. For given any $i = 1, 2, \ldots, r$, $A_i \times T_2$ is covered by a finite subcollection, say $\mathscr{W}_i \subset \mathscr{W}$. Hence $A \times T_2 = \bigcup_{i=1}^{r} (A_i \times T_2)$ is covered by the finite subcollection $\mathscr{W}_1 \cup \mathscr{W}_2 \cup \ldots \cup \mathscr{W}_r$ of $\mathscr{W}$.

STEP 2. We next show that T_1 is locally good, in the sense that for each x in T_1, there is an open set $U(x)$ such that $x \in U(x)$ and $U(x)$ is good.

Proof. Consider a fixed x in T_1. For each y in T_2, $(x, y) \in W(y)$ for some $W(y) \in \mathscr{W}$, since $\mathscr{W}$ covers $T_1 \times T_2$. Now by Proposition 3.5.2 (which is really just the definition of the product topology), there exist sets $U(y)$, $V(y)$ open in T_1, T_2 respectively, such that $(x, y) \in U(y) \times V(y) \subset W(y)$. The collection $\{V(y) : y \in T_2\}$ is an open cover for T_2, so by compactness of T_2 there is a finite subcover, say $\{V(y_1), V(y_2), \ldots, V(y_r)\}$. Let

$$U(x) = U(y_1) \cap U(y_2) \cap \ldots \cap U(y_r)$$

(the idea of this move is illustrated for subsets of $\mathbf{R}^2$ in Fig. 5.2). For each $i = 1, 2, \ldots, r$, $U(x) \times V(y_i) \subset U(y_i) \times V(y_i) \subset W(y_i)$. Hence

$$U(x) \times T_2 = U(x) \times \bigcup_{i=1}^{r} V(y_i) \subset \bigcup_{i=1}^{r} W(y_i),$$

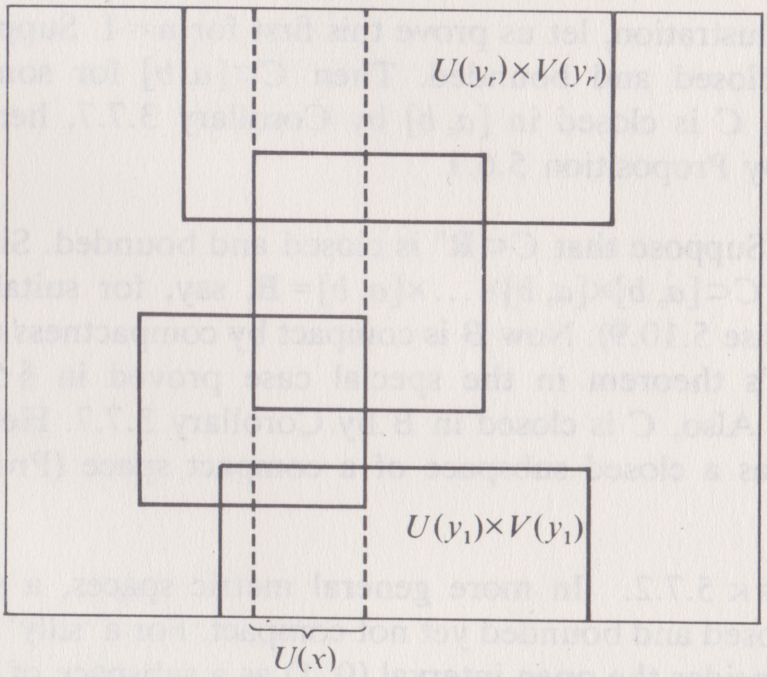


FIG. 5.2

so $U(x)$ is good. Also, $x \in U(x)$ and $U(x)$ is open in T_1, as required.

STEP 3. Finally we use compactness of T_1 to pass from the local to the global. For each x in T_1, let $U(x)$ be a good open set containing x, as provided by Step 2. Then the collection $\{U(x) : x \in T_1\}$ is an open cover for T_1, so by compactness of T_1, there exists a finite subcover, say $\{U(x_1), U(x_2), \ldots, U(x_n)\}$. Since each $U(x_i)$ is good, so is their union, by Step 1. But their union is T_1, so T_1 is good, as required.

REMARK 5.6.3. Theorem 5.6.2 easily extends by induction to the product of any finite number of spaces. It is also true for any (infinite) product of spaces. In this general form it is called Tychonoff's theorem, and it has important applications, for example in modern analysis. Since we are restricting to finite products, we refer to Simmons (1963) for the general case.

5.7 Compact subspaces of $\mathbf{R}^n$

With the help of the results in § 5.6 we may now deduce the following theorem from the compactness of $[a, b]$.

THEOREM 5.7.1 (Heine–Borel). *Any closed bounded subspace C of $\mathbf{R}^n$ is compact.*

As an illustration, let us prove this first for $n = 1$. Suppose that $C \subset \mathbf{R}$ is closed and bounded. Then $C \subset [a, b]$ for some a, b. Moreover, C is closed in $[a, b]$ by Corollary 3.7.7, hence C is compact by Proposition 5.6.1.

Proof. Suppose that $C \subset \mathbf{R}^n$ is closed and bounded. Since C is bounded, $C \subset [a, b] \times [a, b] \times \ldots \times [a, b] = B$, say, for suitable a, b (see Exercise 5.10.9). Now B is compact by compactness of $[a, b]$, Tychonoff's theorem in the special case proved in § 5.6, and induction. Also, C is closed in B by Corollary 3.7.7. Hence C is compact, as a closed subspace of a compact space (Proposition 5.6.1).

REMARK 5.7.2. In more general metric spaces, a subspace may be closed and bounded yet not compact. For a 'silly' example of this, consider the open interval $(0, 1)$ as a subspace of itself. In Chapter 10 we shall see more interesting examples.

5.8 Compactness and uniform continuity

DEFINITION 5.8.1. A map $f: M_1 \to M_2$ of metric spaces with metrics d_1, d_2 is *uniformly continuous on* M_1 if given any $\varepsilon > 0$, there exists $\delta > 0$ such that $d_2(f(x), f(y)) < \varepsilon$ for any x, y in M_1 satisfying $d_1(x, y) < \delta$.

Notice that this is stronger than ordinary continuity of f in that δ can depend on ε but not on x; this is the significance of the adverb *uniformly*. Ordinary continuity of f is a local property, since it says something about the behaviour of f in some neighbourhood of each point of M_1. Uniform continuity is a global property since it says something about the behaviour of f over the whole space M_1. Since compactness allows us to pass from the local to the global, the next proposition is not surprising.

PROPOSITION 5.8.2. *If $f: M_1 \to M_2$ is a continuous map of metric spaces and if M_1 is compact, then f is uniformly continuous on M_1.*

Proof. Let $\varepsilon > 0$. For each x in M_1, there exists $\delta(x) > 0$ such that $d_2(f(x), f(y)) < \frac{1}{2}\varepsilon$ for all y satisfying $d_1(x, y) < 2\delta(x)$, by continuity of f at x (slipping in the multiplier 2 of $\delta(x)$ is more important than usual here). Now the collection $\{B_{\delta(x)}(x) : x \in M_1\}$ is an open cover for M_1. By compactness of M_1 there is a finite subcover $\{B_{\delta(x_1)}(x_1), B_{\delta(x_2)}(x_2), \ldots, B_{\delta(x_r)}(x_r)\}$. Let $\delta = \text{Min}\{\delta(x_1), \delta(x_2), \ldots, \delta(x_r)\}$. Given any x, y in M_1 satisfying $d_1(x, y) < \delta$,

(a) there is some i in $\{1, 2, \ldots, r\}$ such that $d_1(x, x_i) < \delta(x_i)$, and then (b)

$$d_1(y, x_i) \leq d_1(y, x) + d_1(x, x_i) < \delta + \delta(x_i) \leq 2\delta(x_i).$$

Now by (a), $d_2(f(x), f(x_i)) < \frac{1}{2}\varepsilon$, and by (b),

$$d_2(f(y), f(x_i)) < \tfrac{1}{2}\varepsilon,$$

hence

$$d_2(f(x), f(y)) \leq d_2(f(x), f(x_i)) + d_2(f(x_i), f(y)) < \varepsilon,$$

as required.

As a special case of this, any continuous real-valued function on $[a, b]$ is uniformly continuous on $[a, b]$.

REMARK 5.8.3. Compactness of M_1 is not a necessary condition for uniform continuity. For example let M_1 be any noncompact space and let f be the identity map of M_1. Then f is easily seen to be uniformly continuous on M_1.

In terms of uniform continuity, we may define a type of equivalence of metrics which lies between Lipschitz equivalence and topological equivalence (see § 2.4).

DEFINITION 5.8.4. Two metrics d_1, d_2 on a set A are *uniformly equivalent* if the identity map of A is uniformly (d_1, d_2)-continuous and also uniformly (d_2, d_1)-*continuous.*

Exercises 5.10.23–25 explore the comparison with other kinds of equivalence. By Proposition 5.8.2, if $\{A, d_1\}$ is compact and if d_1, d_2 are topologically equivalent metrics then they are uniformly equivalent.

As in § 2.4 there is another closely related definition.

DEFINITION 5.8.5. If $M=\{A, d\}$ and $M'=\{A', d'\}$ are metric spaces then $f: A\to A'$ is a *uniform equivalence* of M with M' if f is a one–one correspondence such that f and f^{-1} are both uniformly continuous.

5.9 An inverse function theorem

THEOREM 5.9.1. *Suppose that $f: T_1\to T_2$ is a continuous one–one correspondence, where T_1 is a compact space and T_2 is a Hausdorff space. Then f is a homeomorphism.*

Proof. Let $g=f^{-1}$. We have to prove that $g: T_2\to T_1$ is continuous. Now for any $V\subset T_1$, $g^{-1}(V)=f(V)$; for $y\in f(V) \Leftrightarrow y=f(x)$ for some x in $V \Leftrightarrow x=g(y)$ is in $V \Leftrightarrow y\in g^{-1}(V)$. We shall prove that if V is closed in T_1 then $g^{-1}(V)$ is closed in T_2. Continuity of g then follows from Proposition 3.7.8.

Suppose that V is closed in T_1. Since T_1 is compact, V is compact (Proposition 5.6.1). Hence $f(V)$ is compact, as the continuous image of a compact space (Proposition 5.5.1), and since T_2 is Hausdorff, $f(V)$ is closed by Proposition 5.4.2. But $f(V)=g^{-1}(V)$, so $g^{-1}(V)$ is closed as required.

COROLLARY 5.9.2. *If $f: T_1\to T_2$ is a continuous injective map, where T_1 is a compact space and T_2 is a Hausdorff space, then f determines a homeomorphism of T_1 onto $f(T_1)$.*

This follows from the theorem and Proposition 3.4.2(b).

EXAMPLE 5.9.3. If $f: [a, b]\to \mathbf{R}$ is a continuous monotonic function then it has a continuous inverse $f^{-1}: f([a, b])\to[a, b]$.

Theorem 5.9.1 is sometimes useful, but in calculus one usually wants to assume more about f and get out differentiability of f^{-1} in the conclusion.

Exercises 5.10

In these exercises, space = topological space.

1. Prove that any space having only a finite number of points is compact.
2. Prove that a discrete space is compact only if it is finite.
3. Given that H, K are compact subspaces of a space T, prove that $H \cup K$ is compact.
4. Given two topologies $\mathcal{T}_1$, $\mathcal{T}_2$ on a set A with $\mathcal{T}_1 \subset \mathcal{T}_2$, prove that if $\{A, \mathcal{T}_2\}$ is compact then so is $\{A, \mathcal{T}_1\}$.
5. Which of the following subsets of $\mathbf{R}$, $\mathbf{R}^2$ are compact?
   (i) $[0, 1)$; (ii) $[0, \infty)$; (iii) $\mathbf{Q} \cap [0, 1]$;
   (iv) $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 = 1\}$;
   (v) $\{(x, y) \in \mathbf{R}^2 : |x| + |y| \leqslant 1\}$;
   (vi) $\{(x, y) \in \mathbf{R}^2 : x^2 + y^2 < 1\}$;
   (vii) $\{(x, y) \in \mathbf{R}^2 : x \geqslant 1, 0 \leqslant y \leqslant 1/x\}$.
6. Prove that the following is a necessary and sufficient condition for a space T to be compact: if $\{V_i : i \in I\}$ is any indexed family of closed subsets of T such that $\bigcap_{i \in J} V_i \neq \varnothing$ for any finite subset $J \subset I$, then $\bigcap_{I} V_i \neq \varnothing$. (Hint: take complements and apply the definition of compactness.)
7. Suppose that for each n in $\mathbf{N}$, V_n is a non-empty closed subset of a compact space T, and $V_n \supset V_{n+1}$. Prove that $\bigcap_{n=1}^{\infty} V_n \neq \varnothing$.
8. Prove that if C, C' are compact subspaces of a Hausdorff space T, then $C \cap C'$ is compact.
9. Prove that any bounded subset of $\mathbf{R}^n$ is contained in a 'cube' $[a, b] \times [a, b] \times \ldots \times [a, b]$ (n factors) for some a, b in $\mathbf{R}$.
10. Show that $\mathbf{R}$ with the Zariski topology of Example 3.1.7 is compact.
11. Prove that if $A \subset \mathbf{R}$ is not compact, then there exists a continuous function $f : A \to \mathbf{R}$ which is not bounded on A. (Hint: consider separately the cases (a) A not bounded, (b) A not closed.)

12. Prove that if $A \subset \mathbf{R}$ is not compact, then there exists a continuous function $f: A \rightarrow \mathbf{R}$ which is bounded but does not attain its bounds on A.

13*. (a) Show that $\mathbf{N}$ may be topologized so that the open sets are $\varnothing$, $\mathbf{N}$ and $\{1, 2, \ldots, n\}$ for every n in $\mathbf{N}$. Call the resulting space T.
   (b) Prove that T is not compact.
   (c) Prove that any continuous map from T to a Hausdorff space is a constant map.
   (d) Deduce that all continuous maps from T to $\mathbf{R}$ are bounded, although T is non-compact.

(Thus in spite of Exercise 11, we cannot take as the definition of compactness of T that all continuous real-valued maps on T are bounded. For further (hard) information on this topic see Hewitt (1948).)

14. Let T be a compact metric space with metric d, and suppose that $f: T \rightarrow T$ is a continuous map such that for every x in T, $f(x) \neq x$. Prove that there exists $\varepsilon > 0$ such that $d(f(x), x) \geqslant \varepsilon$ for all x in T. (Hint: consider $g: T \rightarrow \mathbf{R}$ given by $g(x) = d(f(x), x)$.)

15. Give a precise formulation of the property mentioned in Remark 5.2.3(c). Discuss whether this property may be taken as the definition of compactness.

16. (Bisection argument). Give another proof that $[a, b]$ is compact along the following lines. Let $\mathcal{U}$ be any open cover of $[a, b]$. Assume that there is no finite subcover, and put $c_1 = \frac{1}{2}(a+b)$. Then at least one of $[a, c_1]$, $[c_1, b]$ is not covered by any finite subcollection of $\mathcal{U}$. Let $[a_1, b_1]$ denote one of these intervals which is not so covered, and repeat the argument inductively to get two real sequences (a_n), (b_n) with $a_{n-1} \leqslant a_n \leqslant b_n \leqslant b_{n-1}$ for all n, $b_n - a_n = (b-a)/2^n$, and such that $[a_n, b_n]$ is not covered by any finite subcollection of $\mathcal{U}$. Deduce that (a_n) and (b_n) both converge to some c in $[a, b]$, and get a contradiction by showing that $[a_n, b_n] \subset U$ for some U in $\mathcal{U}$, and sufficiently large n.

17*. Prove that a compact Hausdorff space is regular. Hence prove that it is normal (for definitions see § 4.2).

18. Prove that if $f: M_1 \rightarrow M_2$ and $g: M_2 \rightarrow M_3$ are uniformly

continuous maps of metric spaces then $g \circ f: M_1 \to M_3$ is uniformly continuous.

19. State the negation of the definition of uniform continuity in usable form, and use it to prove that $f: (0, 1) \to \mathbf{R}$ given by $f(x) = 1/x$ is not uniformly continuous on $(0, 1)$.
20. Given that $f: (a, b) \to \mathbf{R}$ is uniformly continuous, show that $\lim_{x \to a+} f(x)$ and $\lim_{x \to b-} f(x)$ exist. (Hint: use the analogue of Theorem 1.2.9, Cauchy's criterion, for limits of functions.)
21. (a) Given that $f: (0, 1) \to \mathbf{R}$ is continuous, monotonic and bounded on $(0, 1)$, show that f is uniformly continuous on $(0, 1)$.

    (b) Is the result still true if the monotonicity assumption is dropped? (Hint for (a): show that $\lim_{x \to 0+} f(x)$ and $\lim_{x \to 1-} f(x)$ exist by analogy with Theorem 1.2.6. Extend f to a continuous function on $[0, 1]$ and use Proposition 5.8.2.)
22. Determine whether $f: A \to \mathbf{R}$ is uniformly continuous on A, when

    (a) $A = (0, 1)$, $f(x) = x$,
    (b) $A = (0, 1)$, $f(x) = 1/(1 - x)$,
    (c) $A = (0, 1)$, $f(x) = \sin 1/x$,
    (d) $A = [1, \infty)$, $f(x) = 1/x$.
23. Show that any two Lipschitz equivalent metrics are uniformly equivalent.
24. Show that for any metric space $M = \{A, d\}$, the metric d_3 given by $d_3(x, y) = d(x, y)/(1 + d(x, y))$ is uniformly equivalent to d, but not necessarily Lipschitz equivalent to d.
25. Show that the metric on $(0, 1) \subset \mathbf{R}$ given by $d(x, y) = |x^{-1} - y^{-1}|$ is topologically equivalent but not uniformly equivalent to the Euclidean metric on $(0, 1)$.
26. Prove that if $A \subset \mathbf{R}$ is not closed, then there exists a continuous function $f: A \to \mathbf{R}$ which is not uniformly continuous.

27*. As in Exercise 4.3.5 let $d(x, K) = \inf_{y \in K} d(x, y)$ for any point x and subset K in a metric space M with metric d. Further let $d(H, K) = \inf_{x \in H} d(x, K)$ for $H, K \subset M$. Prove that if H is compact and K is closed then $d(H, K)=0 \Leftrightarrow H \cap K \neq \varnothing$. Can the hypothesis be relaxed to 'H and K both closed'? (Hint: use the results of Exercise 4.3.5.)

28*. Suppose that $f: T \to T$ is continuous, where T is a compact metric space. Show that there exists a non-empty subset $A \subset T$ such that $f(A)=A$. (Hint: put $T_1=f(T)$, $T_{n+1}=f(T_n)$, $A=\bigcap_{n=1}^{\infty} T_n$ and recall Exercise 7.)

29*. A real-valued map f on a space T is lower semi-continuous (l.s.c.) if for every a in $\mathbf{R}$, $f^{-1}(a, \infty)$ is open in T. Given that T is compact and f is l.s.c. prove that f is bounded below and attains its lower bound on T.

30*. Suppose that $p_1: T_1 \times T_2 \to T_1$ is the projection, where T_2 is compact. Given that $W \subset T_1 \times T_2$ is closed, show that $p_1(W)$ is closed.

31. Suppose that $f: T_1 \to T_2$ is a map from a space T_1 to a compact space T_2, and that its graph G_f is closed in $T_1 \times T_2$. Prove that f is continuous. (Hint: for $V \subset T_2$, prove $f^{-1}(V)= p_1(p_2^{-1}(V) \cap G_f)$. Apply this to closed V, using Exercise 30.)

32*†. Prove that any two norms on $\mathbf{R}^n$ give Lipschitz equivalent metrics.

33. Is there a Hausdorff topology on $[0, 1]$ which is strictly weaker than the Euclidean topology? (Hint: apply Theorem 5.9.1 to the identity map from $[0, 1]$ with its usual topology to $[0, 1]$ with any weaker Hausdorff topology.)

34. Give an example of a continuous injective map $f:(0, 1) \to \mathbf{R}^2$ which does not determine a homeomorphism of $(0, 1)$ onto $f((0, 1))$.

35*. (Alexandroff one-point compactification.) For any space $T=\{A, \mathcal{T}\}$, let $A'=A \cup \{\infty\}$ (where ∞ is any object not in A), and let $\mathcal{T}'$ be the union of $\mathcal{T}$ with all sets of the form $V \cup \{\infty\}$ where $V \subset A$ and $A-V$ is compact and closed in T. Prove that $T'=\{A', \mathcal{T}'\}$ is a compact space containing T as a subspace.

36*. With the notation of Exercise 35, prove that if T is Hausdorff and locally compact (for every x in T there exists an open set U such that $x \in U$ and $Cl(U)$ is compact) then T' is Hausdorff.

6

Connected spaces

6.1 Introduction

INTUITIVELY, a connected space is one which does not fall apart into two or more pieces. But this is not precise, since it does not make clear what 'fall apart' means. We should probably agree that in the real line the subspace $[0, 1]$ is connected whereas $[0, 1] \cup [2, 3]$ is not. But consider for example the subspace $\mathbf{Q}$ of $\mathbf{R}$; or the subspace of the plane consisting of the graph of $y = \sin 1/x$ for $x > 0$ together with the interval on the y-axis joining the points $(0, -1)$ and $(0, 1)$ (see Fig. 6.1 on p. 101). Is either of these spaces connected?

In this chapter we discuss two slightly different formulations of connectedness. The first is the more basic for the study of continuity. As an application we prove the Intermediate Value Theorem 1.4.5. The second formulation (path-connectedness) is possibly closer to intuitive ideas. The two kinds of connectedness are compared in § 6.4.

Both kinds of connectedness are used in the study of functions of a complex variable, at a few crucial stages (see Ahlfors (1966)). In a sense they also form the beginnings of algebraic topology. For example we shall prove that connectedness (and more generally, the number of 'pieces' into which a space falls apart) is a topological invariant. Then one of the easiest ways of showing that two spaces are not homeomorphic, if it works, is to show that one is connected and the other is not (more generally, that they fall apart into different numbers of pieces).

6.2 Connectedness

From the point of view of studying continuity, one meaning which may be assigned to the statement that a space T 'falls apart' into two 'pieces' A and B is this: T is the disjoint union of A and B, and moreover we can define continuous maps on T (to some other topological space) whose values on A 'bear no relation' to their values on B. To make this precise, and to get down to a more specific test, let us define $f: T \to \mathbf{R}$ by

$$f(x) = \begin{cases} 0, & x \in A, \\ 1, & x \in B. \end{cases}$$

Then f fulfils the intuitive property that its values on A 'bear no relation' to its values on B. If f is continuous on T, it is reasonable to think of A and B as mutually independent pieces of T, as far as continuity of real-valued functions on T is concerned. As we shall see in Proposition 6.2.5, this specific test is equivalent to any more comprehensive test along the same lines that one might envisage, of the 'independence' of A and B with regard to continuous maps on T. Moreover, it is just as good to think of f as taking values in the discrete subspace $\{0, 1\}$ of $\mathbf{R}$. We are now within reach of what it means for T not to be connected. By negating this, we get the following definition.

DEFINITION 6.2.1. A topological space T is *connected* if there does not exist any continuous map from T onto a discrete two-point space.

There is an equivalent property which is often taken as the definition. To state it, we need the idea of a partition.

DEFINITION 6.2.2. A *partition* $A \mid B$ of a topological space T is a pair of non-empty subspaces A, B of T such that $A \cup B = T$, $A \cap B = \varnothing$, and both A and B are open in T.

In this situation we say that $A \mid B$ *partitions* T. From the last clause in the definition it follows that B and A are also both closed in T, when $A \mid B$ partitions T. The term partition is overworked: in set theory its appropriate meaning would be just a decomposition into disjoint subsets, and one has to remember that here it has the meaning appropriate to topological spaces.

PROPOSITION 6.2.3. *A topological space T is connected iff it admits no partition.*

Proof. Let D denote $\{0, 1\}$ with the discrete topology. Suppose that $A \mid B$ partitions T. Define $f: T \to D$ by

$$f(x) = \begin{cases} 0, & x \in A, \\ 1, & x \in B. \end{cases}$$

Since A, B are non-empty, f is onto D. The list of all open sets in D

is: $\varnothing$, $\{0\}$, $\{1\}$, D. Now $f^{-1}(\varnothing)=\varnothing$, $f^{-1}(0)=A$, $f^{-1}(1)=B$, $f^{-1}(D)=T$. Thus f is continuous, since each of $\varnothing$, A, B, T is open in T. Hence by definition, T is not connected.

Conversely, if T is not connected, then there exists a continuous onto map $f: T\rightarrow D$, and it is easy to show that $f^{-1}(0)\,|\,f^{-1}(1)$ partitions T.

COROLLARY 6.2.4. *A topological space T is connected iff the only subsets of T which are open and closed in T are $\varnothing$, T.*

Proof. If $A\,|\,B$ partitions T, then A (and B) is an open and closed subset of T which is neither $\varnothing$ nor T. Conversely if there exists an open and closed set A in T other than $\varnothing$, T, then $A\,|\,T-A$ partitions T. The corollary now follows from Proposition 6.2.3.

We shall henceforth use interchangably any of the three criteria for connectedness given in Definition 6.2.1, Proposition 6.2.3, and Corollary 6.2.4.

A non-empty subset S of a topological space T is said to be connected if the space S (with the induced topology) is connected.

The next proposition follows from the special case of Exercise 3.9.33 in which the subsets are disjoint.

PROPOSITION 6.2.5. *Suppose that $A\,|\,B$ partitions a space T, and that $f: T\rightarrow T'$ is any map to another space T' such that the restrictions $f\,|\,A$, $f\,|\,B$ are continuous. Then f is continuous.*

We next aim to show that the connected subspaces of $\mathbf{R}$ are precisely the intervals listed on p. xiii, where a singleton set $\{a\}$ is admitted as an interval $[a, a]$. Before proving this, it is convenient to give a characterization of intervals in terms of 'betweenness'.

LEMMA 6.2.6. *A non-empty subset $S\subset\mathbf{R}$ is an interval iff it satisfies the following property: if x, y in S and z in $\mathbf{R}$ are such that $x<z<y$, then $z\in S$.*

Proof. If S is one of the various intervals, it clearly has the stated property. Conversely, suppose that S has this property. Let

$$a=\inf S \qquad (\text{or } a=-\infty \text{ if } S \text{ is not bounded below}),$$
$$b=\sup S \qquad (\text{or } b=\infty \text{ if } S \text{ is not bounded above}).$$

We shall prove that $(a, b)\subset S\subset[a, b]$, where in order to avoid lists of cases, we make the temporary convention that a bracket adjacent to ±∞ means the same as a parenthesis, and that (a, b) means $\varnothing$ if $a=b$.

First, if $z\in(a, b)$, then $z>a$, so by definition of a there exists x in S with $x<z$. Similarly there exists y in S with $y>z$. Hence by hypothesis $z\in S$. This proves that $(a, b)\subset S$, and the inclusion $S\subset[a, b]$ follows from the definitions of a, b.

But if $(a, b)\subset S\subset[a, b]$ then S can only be one of the various kinds of interval.

Using this lemma we may now prove half of the result we are aiming for.

THEOREM 6.2.7. *No subspace of* **R** *other than an interval is connected.*

Proof. Suppose that $S\subset\mathbf{R}$ is not an interval. Then by Lemma 6.2.6 there exist x, y in S and z in $\mathbf{R}-S$ such that $x<z<y$. Then $S\cap(-\infty, z)\mid S\cap(z, \infty)$ partitions S. For $S\cap(-\infty, z)$ and $S\cap(z, \infty)$ are open in S by definition of the subspace topology. They are non-empty, since they contain x, y respectively. They are clearly disjoint, and their union is S, since $z\notin S$ (it is worth noting that this last clause is the only thing that prevents the same proof from showing that any subset of **R** is not connected).

THEOREM 6.2.8. *Any interval* $I\subset\mathbf{R}$ *is connected.*

Proofs. There are various proofs of this, all ultimately based on the completeness axiom for **R**. We give one proof based directly on the completeness axiom, and another based on compactness of $[a, b]$. A third proof based on compactness of $[a, b]$ will be suggested in Exercise 7.3.9.

The first move is the same for all these proofs. Suppose, to obtain a contradiction, that $A\mid B$ partitions I. Let $a\in A$, $b\in B$, and suppose that $a<b$ (otherwise we may exchange the names of A and B). Since a, $b\in I$ and I is an interval, $[a, b]\subset I$. Let $A'=A\cap[a, b]$, $B'=B\cap[a, b]$. Since A, B are closed in I, by Proposition 3.7.6 A', B' are closed in $[a, b]$. Moreover since $[a, b]$ is closed in **R**, by Corollary 3.7.7, A', B' are closed in **R**. We now give two ways of completing the contradiction.

(1) Since $B'\neq\varnothing$ and B' is bounded below (by a, for example) $b'=\inf B'$ exists by the completeness axiom, and $b'\geqslant a$. Since B' is closed in **R**, by Example 3.7.13 $b'\in B'$. Since $a\in A'$ and $A'\cap B'=\varnothing$, it follows that $b'>a$. Now let $A''=A'\cap[a, b']$.

Arguing as for B', we get that $a''=\sup A''$ exists, that $a''\in A''$ and hence that $a''<b'$. Now $(a'', b')\cap A'=\varnothing$, since otherwise a'' is not an upper bound for A''. Similarly $(a'', b')\cap B'=\varnothing$, otherwise b' is not a lower bound for B'. But $A'\cup B'=[a, b]$ and $(a'', b')\subset[a, b]$. Thus the contradiction is complete.

(2) Since A', B' are closed bounded subsets of $\mathbf{R}$ they are compact (Theorem 5.7.1). Hence $A'\times B'$ is compact (Theorem 5.6.2). Consider the composition

$$A'\times B' \xrightarrow{i} \mathbf{R}\times\mathbf{R} \xrightarrow{d} \mathbf{R},$$

where i is the inclusion and $d(x, y)=|x-y|$. Then d is continuous by a simple special case of Exercise 2.6.17, and since the inclusion i is also continuous, $d\circ i$ is continuous. Hence by Corollary 5.5.4, $d\circ i$ attains its lower bound on $A'\times B'$, say at (a', b') in $A'\times B'$. Since $A'\cap B'=\varnothing$, it follows that $|a'-b'|>0$. We then have $|x-y|\geqslant|a'-b'|$ for all (x, y) in $A'\times B'$. Now what about $c=\frac{1}{2}(a'+b')$? Since a', b' are in $[a, b]$, so is c. But $|c-b'|=\frac{1}{2}|a'-b'|<|a'-b'|$, so (c, b') cannot be in $A'\times B'$, i.e. $c\notin A'$. Similarly $c\notin B'$, and since $A'\cup B'=[a, b]$, the contradiction is again complete.

EXAMPLE 6.2.9. $\mathbf{Q}$ is not connected. For if α is any irrational in $\mathbf{R}$ then $\mathbf{Q}\cap(-\infty, \alpha)\mid\mathbf{Q}\cap(\alpha, \infty)$ partitions $\mathbf{Q}$.

PROPOSITION 6.2.10. *Suppose that T_1 and T_2 are topological spaces with T_1 connected, and that $f: T_1\to T_2$ is continuous. Then $f(T_1)$ is connected.*

Proof. Since the map $f_1: T_1\to f(T_1)$ induced by f is continuous when f is continuous (see Proposition 3.4.2(b)), it is enough to prove the result when f is onto T_2. In this case, if $A\mid B$ partitions T_2, it is easy to see that $f^{-1}(A)\mid f^{-1}(B)$ partitions T_1. This completes the proof. The result is usually stated: a continuous image of a connected space is connected.

COROLLARY 6.2.11. *Connectedness is a topological property.*

COROLLARY 6.2.12. *With the hypotheses of Proposition 6.2.10, the graph G_f of f is connected.*

This follows from the previous corollary since G_f is homeomorphic to T_1 (Proposition 3.6.3).

COROLLARY 6.2.13. *If T is a connected space and $f: T \to \mathbf{R}$ is continuous then $f(T)$ is an interval.*

For the continuous image $f(T)$ of the connected space T is connected (Proposition 6.2.10) and the only connected subspaces of $\mathbf{R}$ are the intervals (including singleton sets) by Theorem 6.2.7.

COROLLARY 6.2.14. *If $f:[a, b] \to \mathbf{R}$ is continuous then f has the I.V.P.*

Proof. Since $[a, b]$ is connected the previous corollary tells us that $f([a, b])$ is an interval. A moment's thought, bearing in mind the betweenness property of intervals, gives the conclusion.

Alternatively, one can prove the Intermediate Value Theorem first and deduce connectedness of $[a, b]$ from it.

We now prove a few standard properties of connectedness.

PROPOSITION 6.2.15. *Suppose that for each i in some indexing set I, A_i is a connected subspace of a space T, and that for all i, j in I, $A_i \cap A_j \neq \varnothing$. Then $A = \bigcup_I A_i$ is connected.*

Proof. Let $f: A \to D$ be continuous, where D is the discrete space $\{0, 1\}$. By definition of connectedness, we have to show that f is not onto D. Since each A_i is connected and $f \mid A_i$ is continuous, $f(A_i) = \{\varepsilon_i\}$ where $\varepsilon_i = 0$ or 1. But for any i, j in I, $A_i \cap A_j \neq \varnothing$, hence $\varepsilon_i = \varepsilon_j$, and f is constant on A as required.

COROLLARY 6.2.16. *If, for each i in some indexing set I, C_i is a connected subspace of a space T, and B is a connected subspace of T such that $B \cap C_i \neq \varnothing$ for each i in I, then $B \cup \bigcup_I C_i$ is connected.*

Proof. Put $A_i = B \cup C_i$ and use the proposition twice.

PROPOSITION 6.2.17. *A product space $T_1 \times T_2$ is connected iff T_1 and T_2 are connected.*

Proof. If $T_1 \times T_2$ is connected then T_1 and T_2, as the continuous images of $T_1 \times T_2$ under the projections p_1 and p_2, are connected.

Suppose that T_1 and T_2 are connected. We may apply Corollary 6.2.16 with $I = T_2$, $C_y = T_1 \times \{y\}$ for each y in T_2, and $B = \{x_0\} \times T_2$ for some fixed x_0 in T_1. For each C_y is homeomorphic to T_1 (Exercise 3.9.19) and hence is connected (Corollary 6.2.11). Likewise, B is connected. Moreover, $C_y \cap B = \{(x_0, y)\}$ is non-empty, and $T_1 \times T_2 = B \cup \bigcup_{T_2} C_y$. Hence by Corollary 6.2.16, $T_1 \times T_2$ is connected.

PROPOSITION 6.2.18. *If H is a connected subspace of a topological space T and if $H \subset K \subset Cl(H)$, then K is connected.*

Proof. Let $f: K \to D$ be any continuous map, where D is the discrete space $\{0, 1\}$. Since H is connected and $f \mid H$ is continuous, $f(H) = \{0\}$ or $\{1\}$. Suppose for definiteness that $f(H) = \{0\}$, and suppose, to get a contradiction, that $f(k) = 1$ for some k in K. Since $\{1\}$ is open in D, $f^{-1}(1)$ is open in K, so since K has the subspace topology, $f^{-1}(1) = K \cap U$ for some U open in T. Now $k \in f^{-1}(1) \subset U$ and $k \in Cl(H)$, so by definition of closure, there exists an x in $H \cap U$. Then since $x \in H$, $f(x) = 0$. But also $f(x) = 1$, since $x \in H \cap U \subset K \cap U = f^{-1}(1)$. This contradiction shows that f cannot be onto D, and K is connected as required.

Alternative proof. By Exercise 3.9.28, the closure $Cl_K(H)$ of H in K is $K \cap Cl_T(H)$. Since $K \subset Cl_T(H)$, this yields $Cl_K(H) = K$. Let $f: K \to D$ be as in the above proof, with $f(H) = \{0\}$. By continuity of f and Exercise 3.9.25, $f(Cl_K(H)) \subset Cl(f(H)) = Cl(\{0\})$. But clearly $Cl(\{0\}) = \{0\}$, so $f(K) = \{0\}$ as required.

Propositions 6.2.15 and 6.2.18 may equally well be proved in the language of partitions.

EXAMPLE 6.2.19. Let G be the graph of $y = \sin 1/x$ for $x > 0$, and let J be the interval joining the points $(0, -1)$ and $(0, 1)$ in $\mathbf{R}^2$. Since the interval $(0, \infty)$ is connected, so is G by Corollary 6.2.12. To show that $G \cup J$ is connected, it is now enough (by Proposition 6.2.18) to show that $J \subset Cl(G)$. Let $p \in J$, say $p = (0, y)$ with $-1 \leq y \leq 1$. We need to show that for any $\varepsilon > 0$, $B_\varepsilon(p) \cap G \neq \varnothing$. Choose n in $\mathbf{N}$ so that $1/2n\pi < \varepsilon$. Since $\sin \frac{1}{2}(4n+1)\pi = 1$ and $\sin \frac{1}{2}(4n+3)\pi = -1$, by the Intermediate Value Theorem $\sin 1/x$ takes on every value between -1 and 1 in the interval $[2/(4n+3)\pi, 2/(4n+1)\pi]$. In particular, $\sin 1/x_0 = y$ for some x_0 in this interval. The distance between the points $(0, y)$ and $(x_0, \sin 1/x_0) = (x_0, y)$ is $x_0 < \varepsilon$, so $(x_0, \sin 1/x_0) \in B_\varepsilon(p) \cap G$, as required.

6.3 Path-connectedness

Intuitively one might say that a subset S of $\mathbf{R}^2$ or $\mathbf{R}^3$ is 'connected' if any two points in S may be joined by a continuous curve lying entirely within S. Indeed, we are more likely to check mentally whether a given (non-pathological) subset of $\mathbf{R}^2$ is 'connected' this way than by referring to Definition 6.2.1. The formal version of this is the definition of path-connectedness, which turns out in general to be slightly different from connectedness (see § 6.4).

Let I denote the interval $[0, 1]\subset\mathbf{R}$.

DEFINITION 6.3.1. Given two points a, b in a topological space T, a *path* from a to b in T is a continuous map $f: I\to T$ such that $f(0)=a$ and $f(1)=b$. We refer to such an f as a path joining a and b.

DEFINITION 6.3.2. A topological space T is *path-connected* if any two points in T can be joined by a path in T.

The next lemma states the intuitively obvious fact that if we go along a path from a to b and then along a path from b to c, the combined operation gives a path from a to c.

LEMMA 6.3.3. *Suppose that f, g: I→T are paths in a space T from a to b and from b to c respectively. Then h: I→T given by*

$$h(x)=\begin{cases} f(2x), & x\in[0, \tfrac{1}{2}],\\ g(2x-1), & x\in[\tfrac{1}{2}, 1],\end{cases}$$

is a path from a to c.

Proof. First, h is well defined, since on $[0, \frac{1}{2}]\cap[\frac{1}{2}, 1]=\{\frac{1}{2}\}$ the two parts of the definition 'fit': $f(1)=b=g(0)$. Now $h\,|\,[0, \frac{1}{2}]=f\circ k$, where $k:[0, \frac{1}{2}]\to[0, 1]$ is given by $k(x)=2x$. Hence $h\,|\,[0, \frac{1}{2}]$ is continuous. Similarly, $h\,|\,[\frac{1}{2}, 1]$ is continuous. By Exercise 3.9.33(i) it follows that h is continuous. Finally,

$$h(0)=f(0)=a, \qquad h(1)=g(1)=c.$$

EXAMPLES 6.3.4. $\mathbf{R}^n$ is path-connected for all $n\geqslant 1$. Any convex subset C of $\mathbf{R}^n$ is path-connected, since (by definition of convexity) any two points in C can be joined by a straight line segment in C. An annulus in $\mathbf{R}^2$, i.e. a set $\{(x, y)\in\mathbf{R}^2 : a\leqslant x^2+y^2\leqslant b\}$ for some $0<a<b$, is path-connected (Exercise 6.6.19).

We leave as exercises the formulation and proofs of analogues for path-connectedness of Proposition 6.2.10, Corollaries 6.2.11, 6.2.12, Proposition 6.2.15, Corollary 6.2.16, and Proposition 6.2.17. The analogue of Proposition 6.2.18 for path-connectedness is false in general, as the next example shows.

EXAMPLE 6.3.5. Let G and J be as in Example 6.2.19. Then $G \cup J$ is not path-connected. This is proved for example on p. 131 of Mendelson (1968). Here is an alternative proof, based on the idea that a continuous path $f: I \rightarrow G \cup J$ which begins at $A = (1/\pi, 0) \in \mathbf{R}^2$ will never arrive at $0 = (0, 0) \in \mathbf{R}^2$, because of the compactness of I.

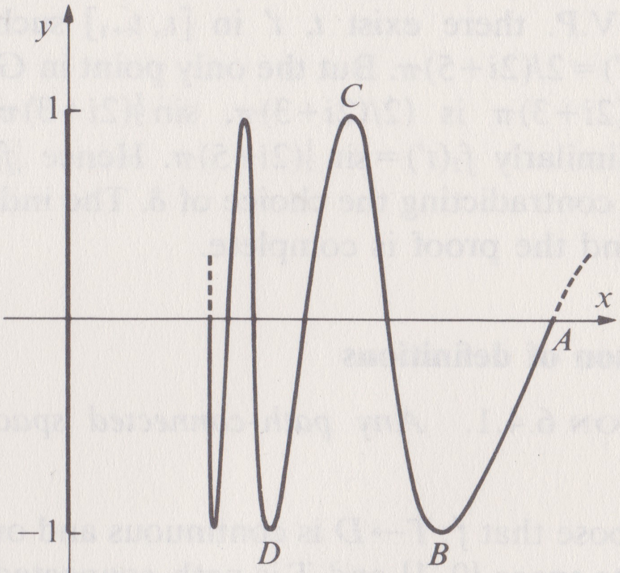


FIG. 6.1

Suppose that $f: I \rightarrow G \cup J$ is continuous and that $f(0) = A$. Let $f_1: I \rightarrow \mathbf{R}$ and $f_2: I \rightarrow \mathbf{R}$ denote $p_1 \circ i \circ f$, $p_2 \circ i \circ f$, where i is the inclusion of $G \cup J$ in $\mathbf{R}^2$ and p_1, p_2 are the projections on the x- and y-axes. Thus f_2 describes the vertical movement of the path and f_1 its horizontal movement. Since f_2 is continuous and I is compact, f_2 is uniformly continuous on I by Proposition 5.8.2. Let $\delta > 0$ be such that $|f_2(t) - f_2(t')| < 2$ for any t, t' in I satisfying $|t - t'| < \delta$. Let $0 = t_0 < t_1 < \ldots < t_n = 1$ be such that $t_i - t_{i-1} < \delta$ for $i = 1, 2, \ldots, n$. The idea now is as follows. As t goes from t_0 to t_1, $f(t)$, which starts at A, cannot get as far as C in Fig. 6.1, since it would have to pass through B on the way; but then $f(t) = B$ and $f(t') = C$ for some t, t' with $|t - t'| < \delta$, and since B and C are a vertical distance of 2 apart, $|f_2(t) - f_2(t')| = 2$, contradicting the choice of δ. Similarly, as t goes from t_1 to t_2, $f(t)$ cannot get as far as D, and so on; as t goes from 0 to 1, $f(t)$ cannot traverse more than n ups and downs in all.

Formally, we shall prove inductively that

$$f_1(t_i) > 2/(2i+3)\pi \qquad \text{for} \qquad i = 0, 1, \ldots, n.$$

This is sufficient, for then $f_1(1) > 2/(2n+3)\pi > 0$, so $f(1) \neq (0, 0)$.

Since $f(t_0) = f(0) = (1/\pi, 0)$, $f_1(t_0) = 1/\pi > 2/3\pi$. Suppose inductively that $f_1(t_i) > 2/(2i+3)\pi$ for some $i \geqslant 0$, and suppose also, to get a contradiction, that $f_1(t_{i+1}) \leqslant 2/(2i+5)\pi$. Then since f_1 is continuous, by the I.V.P. there exist t, t' in $[t_i, t_{i+1}]$ such that $f_1(t) = 2/(2i+3)\pi$, $f_1(t') = 2/(2i+5)\pi$. But the only point in $G \cup J$ with first coordinate $2/(2i+3)\pi$ is $(2/(2i+3)\pi, \sin\frac{1}{2}(2i+3)\pi)$, so $f_2(t) = \sin\frac{1}{2}(2i+3)\pi$. Similarly $f_2(t') = \sin\frac{1}{2}(2i+5)\pi$. Hence $|f_2(t) - f_2(t')| = 2$ while $|t - t'| < \delta$, contradicting the choice of δ. The inductive step is now proved, and the proof is complete.

6.4 Comparison of definitions

PROPOSITION 6.4.1. *Any path-connected space T is connected.*

Proof. Suppose that $f: T \to D$ is continuous and onto D, where D is the discrete space $\{0, 1\}$ and T is path-connected. Let $x, y \in T$ with $f(x) = 0$, $f(y) = 1$. Let $g: I \to T$ be a path from x to y. Then $f \circ g: I \to D$ is continuous and onto D, contradicting the connectedness of I. Hence T must be connected.

Examples 6.2.19 and 6.3.5 show that the converse to Proposition 6.4.1 is false in general. However, it is true for a large class of 'reasonable' spaces. We give one example of this.

PROPOSITION 6.4.2. *A connected open subset U of $\mathbf{R}^n$ is path-connected.*

Proof. Let $a \in U$, let $H \subset U$ be the subset of points which can be joined to a by a path in U, and let $K = U - H$. We shall prove that if K were non-empty then $H \mid K$ would partition U, contradicting its connectedness.

First we show that H is open (in $\mathbf{R}^n$ or in U, equivalently since U is open in $\mathbf{R}^n$). For let $x \in H$. Then for some $\varepsilon > 0$, $B_\varepsilon(x) \subset U$. Given any y in $B_\varepsilon(x)$, there is a (straight line) path g in $B_\varepsilon(x) \subset U$, joining x to y. But since x is in H, there is a path f in U joining a to x. The path h (as in Lemma 6.3.3) formed by traversing f and then g is a path in U from a to y. Hence $y \in H$, so $B_\varepsilon(x) \subset H$ and H is open.

By a similar argument, K is open. (If $x \in K$, then $B_\varepsilon(x) \subset U$ for some $\varepsilon > 0$, and if any point in $B_\varepsilon(x)$ could be joined to a by a path in U then so could x.) It is clear that $H \cap K = \varnothing$ and $H \cup K = U$, so by connectedness of U we conclude that $K = \varnothing$, and $H = U$ as required.

The same proof shows that any two points in a connected open subset of $\mathbf{R}^n$ may be joined by a polygonal path (a finite union of straight line segments).

6.5 Components

Given a topological space T, let us define an equivalence relation on T as follows: $x \sim y$ iff x and y are both contained in some connected subspace of T. This relation is clearly reflexive and symmetric; transitivity follows from Proposition 6.2.15. The resulting equivalence classes are called the *components* of T. Alternatively we may define a component of T to be a maximal connected subset of T. (The equivalence of the definitions is left as Exercise 6.6.18.)

EXAMPLES 6.5.1. The components of $[0, 1] \cup [2, 3]$ are $[0, 1]$ and $[2, 3]$. The components of $\mathbf{Q}$ are singleton sets. For if $x, y \in \mathbf{Q}$ with $x \neq y$, then there is no connected subspace S of $\mathbf{Q}$ containing x and y. (If $x, y \in S \subset \mathbf{Q}$, let α be any irrational between x and y. Then $S \cap (-\infty, \alpha) \mid S \cap (\alpha, \infty)$ partitions S, so S is not connected.)

A topological space in which the components are all singleton sets is called *totally disconnected.* It is important to note that this does not imply that the space is discrete, although any discrete space is totally disconnected. For example in $\mathbf{Q}$, singleton sets are not open, so $\mathbf{Q}$ is not discrete. (Any non-empty open subset of $\mathbf{Q}$ is of the form $U \cap \mathbf{Q}$ for some non-empty open subset U of $\mathbf{R}$. Then U contains an open interval surrounding each of its points, so U contains infinitely many points of $\mathbf{Q}$.)

For any space T, let $\mathscr{H}(T)$ denote the set of its components. For any continuous map $f: T_1 \to T_2$ we may define a corresponding set map $\mathscr{H}(f): \mathscr{H}(T_1) \to \mathscr{H}(T_2)$ as follows: if C_x is the component of T_1 containing x, let $\mathscr{H}(f)(C_x)$ be the component $C_{f(x)}$ of T_2 containing $f(x)$. Then $\mathscr{H}(f)$ is well-defined, because if $C_x = C_y$ then $x, y \in C$ for some connected subset C of T_1, so $f(x), f(y) \in f(C)$, and $f(C)$, as the continuous image of a connected space, is connected. Hence $C_{f(y)} = C_{f(y)}$.

PROPOSITION 6.5.2. *Given continuous maps $f: T_1 \to T_2$ and $g: T_2 \to T_3$, the following diagram commutes:*

$$\begin{tikzcd}
\mathcal{H}(T_1) \arrow[rr, "\mathcal{H}(f)"] \arrow[dr, "\mathcal{H}(g\circ f)"'] & & \mathcal{H}(T_2) \arrow[dl, "\mathcal{H}(g)"] \\
& \mathcal{H}(T_3) &
\end{tikzcd}$$

Proof. For any x in T_1,

$$\mathcal{H}(g)\circ\mathcal{H}(f)(C_x) = \mathcal{H}(g)(C_{f(x)}) = C_{g(f(x))} = C_{g\circ f(x)} = \mathcal{H}(g\circ f)(C_x).$$

Hence $\mathcal{H}(g)\circ\mathcal{H}(f) = \mathcal{H}(g\circ f)$.

PROPOSITION 6.5.3. *If $1: T \to T$ denotes the identity map of a space T, then $\mathcal{H}(1): \mathcal{H}(T) \to \mathcal{H}(T)$ is the identity map 1 of $\mathcal{H}(T)$.*

The proof is immediate.

†Readers who know categories and functors will recognize that we have just proved that $\mathcal{H}$ is a covariant functor from the category of topological spaces and continuous maps to the category of sets and set maps, and they will not be surprised by the next deduction.†

THEOREM 6.5.4. *If $f: T_1 \to T_2$ is a homeomorphism then $\mathcal{H}(f): \mathcal{H}(T_1) \to \mathcal{H}(T_2)$ is a one–one correspondence.*

Proof. By Propositions 6.5.2 and 6.5.3,

$$\mathcal{H}(f)\circ\mathcal{H}(f^{-1}) = \mathcal{H}(f\circ f^{-1}) = \mathcal{H}(1) = 1,$$

and likewise $\mathcal{H}(f^{-1})\circ\mathcal{H}(f) = 1$. It follows that $\mathcal{H}(f)$ is a one–one correspondence.

In particular if the number $|\mathcal{H}(T)|$ of components of a space T is finite, then for any space S which is homeomorphic to T we get $|\mathcal{H}(S)| = |\mathcal{H}(T)|$. (The proof needed in Example 3.6.2(d) may now be completed.) The number of components of a space is a simple example of the kind of topological invariants used in algebraic topology.

The components of a space T are always closed in T, by Proposition 6.2.18, but they may not be open in T (as Example 6.5.1 shows for $T = \mathbf{Q}$).

DEFINITION 6.5.5. A space T is *locally connected* if given any $x \in U \subset T$, where U is open in T, there exists a connected open set V such that $x \in V \subset U$. We mention without proof that the components of a locally connected space T are open in T, and that the space $G \cup J$ of Examples 6.2.19 and 6.3.5 is not locally connected.

Path-components may be defined similarly to components, and the analogues of Propositions 6.5.2, 6.5.3 and Theorem 6.5.4 all go through.

Exercises 6.6

1. Prove that if $f: T \to D$ is a continuous map from a space T onto a discrete space $D = \{0, 1\}$, then $f^{-1}(0) \mid f^{-1}(1)$ partitions T.
2. Show that a space T is connected iff for every $H \subset T$ such that $H \neq \varnothing$, T, the boundary $b(H)$ is non-empty. (Hint: see Exercise 3.9.30(ii).)
3. Which of the following subsets of $\mathbf{R}^2$ are (a) connected and (b) path-connected?
   (i) the mutually tangent open discs $B_1(1, 0) \cup B_1(-1, 0)$,
   (ii) $Cl(B_1(1, 0)) \cup Cl(B_1(-1, 0))$,
   (iii) $Cl(B_1(1, 0)) \cup B_1(-1, 0)$,
   (iv) the rational comb
   $$\{(q, y) \in \mathbf{R}^2 : q \in \mathbf{Q},\ y \in [0, 1]\} \cup (\mathbf{R} \times \{1\}),$$
   (v) the set of all points with at least one coordinate in $\mathbf{Q}$.
4. Is Example 3.1.7 connected?
5. If $\mathcal{T}_1 \subset \mathcal{T}_2$ are two topologies on a set A, does either of the following implications hold: $\{A, \mathcal{T}_1\}$ connected $\Rightarrow$ $(\Leftarrow)$ $\{A, \mathcal{T}_2\}$ connected?
6. A map $f: T \to \mathbf{R}$ on a topological space T is *locally constant* if for each x in T there is an open set U with $x \in U$ and $f \mid U$ a constant function. Prove that if T is connected then any locally constant map f on T is a constant map. Does this still hold when $\mathbf{R}$ is replaced by an arbitrary topological space?
7. If T is a connected space containing more than one point, and if $\{x\}$ is closed for every x in T, show that the number of points in T is infinite.

8. Deduce from the Intermediate Value Theorem that $[0, 1]$ is connected.

Exercises 9–15 can all be done using the Intermediate Value Theorem.

9. Show that any polynomial equation of odd degree with real coefficients has at least one real root.

10. Prove that any continuous function $f:[a, b]\to[a, b]$ has a fixed point, that is, a point x such that $f(x)=x$. (Hint: consider the function $g:[a, b]\to\mathbf{R}$ given by $g(x)=f(x)-x$.)

11. (a) Prove that if T is a connected space and if there exists a non-constant real-valued continuous function on T, then T is uncountably infinite.
    (b) Prove that a connected metric space M containing more than one point contains uncountably many points.

12. Prove that there is no continuous injective map $f:\mathbf{R}^2\to\mathbf{R}$.

13. Suppose that $f:[0, 1]\to\mathbf{R}$ is continuous with $f(0)=f(1)$, and that $n\in\mathbf{N}$, $n\geqslant 2$. Show that there exists an x in $[0, 1]$ such that $f(x)=f(x+1/n)$. (Hint: let $g(x)=f(x)-f(x+1/n)$, and show that $g(0)+g(1/n)+\ldots+g((n-1)/n)=0$.)

14*. Let $f:\mathbf{R}\to\mathbf{R}$ be a function such that for each x in $\mathbf{R}$, $f^{-1}(x)$ contains exactly two points. Prove that f is not continuous.

15*. Prove that there is no continuous function $f:[0, 1]\to\mathbf{R}$ satisfying $x\in\mathbf{Q}\Leftrightarrow f(x)\notin\mathbf{Q}$.

16*. Let $\mathbf{S}\subset\mathbf{R}^2$ denote the circle $\{(x_1, x_2)\in\mathbf{R}^2 : x_1^2+x_2^2=1\}$, and let $f:\mathbf{S}\to\mathbf{R}$ be a continuous map. Show that there exists a point $x=(x_1, x_2)$ in $\mathbf{S}$ such that $f(x)=f(-x)$, where $-x=(-x_1, -x_2)$. (Hint: consider the function $g:\mathbf{S}\to\mathbf{R}$ given by $g(x)=f(x)-f(-x)$.)

17. Prove that any discrete space is totally disconnected.

18. Prove that the components of a space are its maximal connected subsets.

19. Prove that the annulus in Example 6.3.4 is path-connected.

20*. Prove that if H, K are non-empty closed subsets of a space T such that $H\cup K$ and $H\cap K$ are connected then H and K are connected.

21*. Give an example of closed connected subsets $C_n \subset \mathbf{R}^2$ such that $C_n \supset C_{n+1}$ for all n in $\mathbf{N}$, but $\bigcap_{n=1}^{\infty} C_n$ is not connected.

22*. Let C_n be a compact connected subspace of a Hausdorff space such that $C_n \supset C_{n+1}$ for all n in $\mathbf{N}$. Prove that $\bigcap_{n=1}^{\infty} C_n$ is connected.

23. By considering a closed disc in $\mathbf{R}^2$ and the union of two disjoint closed discs in $\mathbf{R}^2$, show that there exist non-homeomorphic compact Hausdorff spaces T_1, T_2 with continuous injective maps $f: T_1 \to T_2$ and $g: T_2 \to T_1$.

7

Compactness again: convergence in metric spaces

7.1 Introduction

We saw in Chapter 4 how convergence may be defined for sequences in topological spaces. The study of convergence leads to an alternative approach to compactness, giving rise to the related concept of sequential compactness. We shall consider sequential compactness in metric spaces only, and show that in this case the two kinds of compactness are equivalent. Some motivation for studying convergence in metric spaces was mentioned at the beginning of § 1.2.

7.2 Sequential compactness

Definitions 7.2.1. A subspace C of a metric space M is *sequentially compact in itself* (*in* M) if any sequence in C has a subsequence converging to a point of C (of M).

If C is sequentially compact in itself then it is sequentially compact in M, by the definitions. If C is sequentially compact in M then C is sequentially compact in itself iff C is closed in M, by the next lemma.

Lemma 7.2.2. *Let C be a subset of a metric space M and let $x \in M$. Then $x \in Cl(C)$ iff there exists a sequence (x_n) in C converging to x.*

Proof. First suppose that (x_n) is a sequence in C converging to x. Given $\varepsilon > 0$, there exists N such that $x_n \in B_\varepsilon(x)$ for all $n \geq N$. Since $x_n \in C$ for all n, this means that for every $\varepsilon > 0$, $B_\varepsilon(x) \cap C \neq \varnothing$. Hence $x \in Cl(C)$.

Conversely suppose that $x \in Cl(C)$. By definition of closure, there exists a point x_n in $C \cap B_{1/n}(x)$ for every n in $\mathbf{N}$, and clearly (x_n) converges to x.

Working from Definition 7.2.1, we could now prove that if C is sequentially compact in itself then it is bounded as well as closed, and that a closed bounded subspace of $\mathbf{R}^n$ is sequentially compact in itself. However, these results will follow from Chapter 5 once we have proved that for metric spaces sequential compactness coincides with compactness.

DEFINITION 7.2.3. A sequence (x_n) in a metric space M with metric d is a *Cauchy sequence* if given $\varepsilon>0$, there exists N such that $d(x_n, x_m)<\varepsilon$ for all $m, n\geqslant N$.

Cauchy sequences will be studied again in later chapters. All we need at present is the following, whose proof is as for the easy part of Theorem 1.2.9.

PROPOSITION 7.2.4. *Any convergent sequence in a metric space is a Cauchy sequence.*

Next we prove an intuitively obvious lemma about limit points, which is a more delicate version of one half of Lemma 7.2.2.

LEMMA 7.2.5. *Suppose that (x_n) is a sequence in a metric space and that x is a limit point of the set of members of (x_n), $S=\{x_n : n\in\mathbf{N}\}$. Then (x_n) has a subsequence converging to x.*

Proof. The proof is a little clumsy, to ensure that we really get a subsequence. First, S is infinite by Exercise 3.9.22. We may assume that x_n never equals x. For otherwise, delete all the occurrences of x from (x_n): the result is still a sequence since S is infinite, and its set of members $S - \{x\}$ still has x as a limit point.

Since x is a limit point of S, there exists an integer, $n(1)$ say, such that $x_{n(1)}\in B_1(x)$. Suppose inductively that integers $n(1) < n(2) < \ldots < n(k)$ have been chosen so that $x_{n(i)}\in B_{l/i}(x)$ for $i = 1,2,\ldots, k$. Put

$$\varepsilon=\operatorname{Min}\{1/(k+1), d(x_1, x), d(x_2, x),\ldots, d(x_{n(k)}, x)\}.$$

Since $x_n\neq x$ for any n, $\varepsilon>0$. Since x is a limit point of S, there exists an integer $n(k+1)$ such that $x_{n(k+1)}\in B_\varepsilon(x)$. Now ε has been chosen to force $n(k+1)>n(k)$, since $d(x_i, x)\geqslant\varepsilon$ for all $i\leqslant n(k)$. Also, $x_{n(k+1)}\in B_{1/(k+1)}(x)$. This completes the inductive step in the construction of a subsequence, which converges to x since for all k_0 in $\mathbf{N}$ and all $k\geqslant k_0$ we have $d(x_{n(k)}, x)<1/k\leqslant 1/k_0$.

We now prove half of the equivalence of definitions of compactness.

PROPOSITION 7.2.6. *If a subspace C of a metric space M is compact then it is sequentially compact in itself.*

Proof. Suppose that C is compact (in the sense of Chapter 5). Let (x_n) be a sequence in C, and let S denote the set of members of (x_n). Thus $S \subset C$, and S may be finite or infinite.

First suppose that S is finite. Then at least one point, say x, of S must be repeated infinitely often in the sequence (x_n), and its occurrences form a subsequence converging to x, which is a point of C.

Now suppose that S is infinite. It is enough, by Lemma 7.2.5, to show that S has a limit point in C. Suppose, to get a contradiction, that S has no limit point in C. Then for each x in C, there exists $\varepsilon(x)>0$ such that $B_{\varepsilon(x)}(x)$ contains no point of S other than x. In other words, $B_{\varepsilon(x)}(x) \cap S = \varnothing$ or $\{x\}$. Now by compactness of C, the open cover $\{B_{\varepsilon(x)}(x) : x \in C\}$ has a finite subcover. But the union of any finite subcollection of the $B_{\varepsilon(x)}(x)$ contains only finitely many points of S, since each $B_{\varepsilon(x)}(x)$ contains at most one point of S. Hence no finite subcollection covers S, let alone C. This contradiction shows that S must have a limit point in C, as required.

COROLLARY 7.2.7. *Any bounded sequence in $\mathbf{R}^n$ has a convergent subsequence.*

Proof. The set S of members of the sequence is bounded, hence $Cl(S)$ is a closed, bounded (cf. Exercise 3.9.32) subset of $\mathbf{R}^n$, and the result follows by Theorem 5.7.1 and Proposition 7.2.6.

There are many different routes to proving Corollary 7.2.7. Another is suggested in Exercise 7.3.6.

We now prepare for the converse to Proposition 7.2.6.

DEFINITION 7.2.8. Given $\varepsilon>0$, an *ε-net* for a metric space M is a subset $S \subset M$ such that $M \subset \bigcup_{x \in S} B_\varepsilon(x)$.

Intuitively, when $M \subset \mathbf{R}^2$ we think of the points in S as the crossings of a net. The definition requires that each point of M is within ε of at least one crossing. Of course, we can always get an ε-net for any $\varepsilon>0$ by taking $S=M$. The interesting ε-nets are the finite ones. (Accordingly, finiteness is sometimes taken to be part of the definition.)

PROPOSITION 7.2.9. *If a metric space M is sequentially compact, then given any $\varepsilon>0$, there exists a finite ε-net for M.*

Proof. Let d be the metric for M. Suppose that for some $\varepsilon>0$, there is no finite ε-net for M. We shall get a contradiction by constructing a sequence (x_n) in M with no convergent subsequence.

Let x_1 be an arbitrary point in M, and suppose inductively that $x_1, x_2, \ldots, x_r$ have been chosen in M such that $d(x_i, x_j)\geqslant\varepsilon$ for all $i, j\leqslant r$, $i\neq j$. By hypothesis, $\{x_1, x_2, \ldots, x_r\}$ is not an ε-net, so there exists a point in M, which we call x_{r+1}, such that $d(x_{r+1}, x_i)\geqslant\varepsilon$ for $1\leqslant i\leqslant r$. This completes the inductive step in the construction of a sequence (x_n) such that $d(x_m, x_n)\geqslant\varepsilon$ for all $m\neq n$. This sequence has clearly no Cauchy subsequence. Hence by Proposition 7.2.4 it has no convergent subsequence either.

The property which a sequentially compact metric space is asserted to have in this proposition is important enough to have two names; we shall study it again in Chapter 10.

DEFINITION 7.2.10. A metric space is *totally bounded* or *precompact* if it has a finite ε-net for every $\varepsilon>0$.

DEFINITION 7.2.11. Given an open cover $\mathcal{U}$ of a metric space M, a fixed real number $\varepsilon>0$ is called a *Lebesgue number* for $\mathcal{U}$ if for any x in M, there exists a set $U(x)$ in $\mathcal{U}$ such that $B_\varepsilon(x)\subset U(x)$.

If ε is a Lebesgue number for $\mathcal{U}$ then so is any ε' satisfying $0<\varepsilon'\leqslant\varepsilon$. There may or may not be any Lebesgue number for a given open cover. For example let $M=(0, 1)\subset\mathbf{R}$, and let $\mathcal{U}=\{(1/n, 1) : n\geqslant 2\}$. Then for any $\varepsilon>0$, take $n>1/\varepsilon$ and $x=1/n$. There is no $(1/m, 1)$ such that $B_\varepsilon(x)\subset(1/m, 1)$, since $B_\varepsilon(x)=B_\varepsilon(1/n)=(0, 1/n+\varepsilon)$.

LEMMA 7.2.12. *There exists a Lebesgue number for any open cover of a sequentially compact metric space M.*

Proof. Suppose that $\mathcal{U}$ is an open cover for which no Lebesgue number exists. Then for any integer n, there exists some point of M, say x_n, such that $B_{1/n}(x_n)\subset U$ is false for every U in $\mathcal{U}$, since otherwise $1/n$ would be a Lebesgue number for $\mathcal{U}$. By sequential compactness, (x_n) has a subsequence, say $(x_{n(r)})$, converging to a point x of M. Since $\mathcal{U}$ covers M, $x\in U_0$ for some U_0 in $\mathcal{U}$. Since U_0 is open, $B_{2/m}(x)\subset U_0$ for some integer m. Now $B_{1/m}(x)$ contains $x_{n(r)}$

for all $r \geqslant R$, say. Choose $r \geqslant R$ so that $n(r) \geqslant m$, and write $s = n(r)$. Then $B_{1/s}(x_s) \subset B_{2/m}(x)$, since (draw a picture in $\mathbf{R}^2$)

$$d(x_s, y) < 1/s \Rightarrow d(x, y) \leqslant d(x, x_s) + d(x_s, y) < 1/m + 1/s \leqslant 2/m.$$

Hence $B_{1/s}(x_s) \subset U_0$, contradicting the choice of x_s. This contradiction shows that there exists a Lebesgue number for $\mathcal{U}$.

THEOREM 7.2.13. *Any sequentially compact metric space is compact.*

Proof. Let $\mathcal{U}$ be an arbitrary open cover of a sequentially compact metric space M. By Lemma 7.2.12 there exists a Lebesgue number ε for $\mathcal{U}$. By Proposition 7.2.9 there exists a finite ε-net $\{x_1, x_2, \dots, x_n\}$ for M, where ε is this same Lebesgue number. Then $B_\varepsilon(x_i)$ is contained in some set, say U_i, in $\mathcal{U}$ by definition of Lebesgue number. Since $M \subset \bigcup_{i=1}^{n} B_\varepsilon(x_i) \subset \bigcup_{i=1}^{n} U_i$, we have a finite subcover $\{U_1, U_2, \dots, U_n\}$ of $\mathcal{U}$ for M.

This completes the proof of equivalence of the two definitions of compactness for a metric space. From now on, when we refer to compactness of a metric space, we shall use the definition which is more convenient at the time.

At this stage our only examples of compact metric spaces are finite spaces and closed bounded subspaces of $\mathbf{R}^n$. This situation will be remedied in Chapter 10.

Exercises 7.3

1. Prove that any convergent sequence in a metric space is a Cauchy sequence.
2. Prove that any Cauchy sequence in a metric space is bounded.
3. Give direct proofs (not using Theorem 7.2.13) that
   (a) a sequentially compact metric space is bounded,
   (b) a closed subspace of a sequentially compact metric space is sequentially compact,
   (c) the product of two sequentially compact metric spaces is sequentially compact.
4. Prove that the following is a necessary and sufficient condition for a map $f: M_1 \to M_2$ of metric spaces to be continuous at a: if (x_n) is any sequence in M_1 converging to a, then $(f(x_n))$ converges to $f(a)$.

5. Given that d, d' are topologically equivalent metrics on a set A, show that a sequence (x_n) in A converges to a in $\{A, d\}$ iff it converges to a in $\{A, d'\}$.
6. Give a direct proof that a closed bounded subspace of $\mathbf{R}^n$ is sequentially compact. (Hint: use a bisection argument in $\mathbf{R}$ and then Exercise 3(b), (c).)
7. Give an example of a metric space which is precompact but not compact.
8. Give another proof that a continuous map $f: M_1 \to M_2$ of metric spaces is uniformly continuous if M_1 is compact, along the following lines: let $\varepsilon > 0$, and let $\delta > 0$ be a Lebesgue number for the open cover $\{f^{-1}(B_{\frac{1}{2}\varepsilon}(y)) : y \in M_2\}$ of M_1. Show that $d_1(x, x') < \delta \Rightarrow d_2(f(x), f(x')) < \varepsilon$, where d_1, d_2 are the metrics on M_1, M_2.
9. Give another proof that $[a, b]$ is connected, along the following lines: suppose that $A \mid B$ partitions $[a, b]$ where $a \in A$. Since A, B are open in $[a, b]$, for each x in $[a, b]$ there is some $B_{\delta(x)}(x)$ entirely contained in either A or B. Let ε be a Lebesgue number for $\{B_{\delta(x)}(x) : x \in [a, b]\}$, and let $a = a_0 < a_1 < \ldots < a_n = b$ be such that $a_i - a_{i-1} < \varepsilon$ for $i = 1, 2, \ldots, n$. Deduce that $[a, b] \subset A$.

10*. Let C denote the subspace of Hilbert space ℓ^2 consisting of all sequences (x_n) satisfying $|x_n| \leqslant 1/n$ for every n. Prove that C is compact.

11*. Let M be a compact metric space and suppose that for every n in $\mathbf{N}$, V_n is a closed subset of M such that $V_n \supset V_{n+1}$. Prove that

$$\operatorname{diam}\left(\bigcap_{n=1}^{\infty} V_n\right) = \inf\{\operatorname{diam} V_n : n \in \mathbf{N}\}.$$

12*. (a) Let M be a compact metric space with metric d, and suppose that $f: M \to M$ satisfies $d(f(x), f(y)) = d(x, y)$ for all x, y in M. Prove that f is onto M. (So f is an isometry.)

(b) Let M_1, M_2 be compact metric spaces with metrics d_1, d_2 and suppose that $f: M_1 \to M_2$, $g: M_2 \to M_1$ satisfy $d_2(f(x), f(x')) = d_1(x, x')$, $d_1(g(y), g(y')) = d_2(y, y')$ for all x, x' in M_1, and all y, y' in M_2. Prove that f and g are both isometries. (Contrast this with Exercise 6.6.23.)

(c) Construct a function $f: (0, \infty) \to (0, \infty)$ which is not onto $(0, \infty)$, but satisfies $|f(x) - f(y)| = |x - y|$ for all x, y in $(0, \infty)$.

8

Uniform convergence

8.1 Introduction

In order to illustrate the scope of compactness (Chapters 5, 7) and completeness (Chapter 9) it is desirable to consider an application outside of Euclidean spaces. The most easily accessible example concerns function spaces with the sup metric. To prepare for further study of these function spaces, in this chapter we introduce uniform convergence in the particular context of real-valued functions of a real variable, pointing out the relation with convergence in a function space, and at the end mentioning generalizations.

The study of uniform convergence may be motivated as follows. Many particular functions in analysis are defined by means of sequences (or series). Likewise, existence proofs for solutions of differential and other equations often produce the solution as the limit of a sequence of functions. In such cases, we want to know whether the limit is continuous. (Similarly, we want to know whether we can differentiate or integrate it by differentiating or integrating each term in the sequence and taking the limit.) Uniform convergence contributes to sufficient, though not usually necessary, conditions for ensuring that the limit function is well-behaved if the terms in the sequence are well-behaved. More generally, it often provides sufficient conditions for permuting the order in which limiting processes are carried out, in a sense which will be illustrated in Remark 8.4.2(a) below.

8.2 Definition and examples

We shall be concerned initially with real-valued functions defined on some subset $D \subset \mathbf{R}$. In applications, D is usually an interval. Suppose that (f_n) is a sequence of real-valued functions, and that the domain of each f_n contains D. If we fix attention on a particular point x in D, the values of the f_n at x give a sequence

$(f_n(x))$ in $\mathbf{R}$. Suppose that for each x in D, $(f_n(x))$ converges. Then we may define a new function $f: D \to \mathbf{R}$ by putting $f(x) = \lim_{n\to\infty} f_n(x)$. A slightly different viewpoint is to suppose given functions f_n and f from D to $\mathbf{R}$, and to make the following definition.

DEFINITION 8.2.1. The sequence (f_n) *converges to f pointwise on D* if for each x in D, the real number sequence $(f_n(x))$ converges to $f(x)$.

Pointwise convergence has curious features which may sometimes be disadvantageous. For example, we cannot deduce continuity of f from continuity of all the f_n.

EXAMPLE 8.2.2. Let $f_n(x) = x^n$, $x \in [0, 1]$. Writing

$$f(x) = \lim_{n\to\infty} f_n(x),$$

we get

$$f(x) = \begin{cases} 0, & x \in [0, 1), \\ 1, & x = 1. \end{cases}$$

Thus the limit function is discontinuous at $x = 1$, although all the functions in the sequence are continuous. (The reader can gain insight on this by drawing the graphs of f_n for $n = 1, 2, 3, 4$.)

In certain situations (e.g. in Fourier series) this kind of thing must happen. But often we would like the limit function to be continuous. Uniform convergence helps to ensure this, as we shall see in Theorem 8.4.1.

DEFINITION 8.2.3. A sequence (f_n) of real-valued functions defined on $D \subset \mathbf{R}$ *converges to a function f uniformly on D* if given $\varepsilon > 0$, there exists N such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \geqslant N$ and all x in D.

If this is fulfilled, we write '$f_n \to f$ uniformly on D' and call f the 'uniform limit of f_n'. Obviously, if $f_n \to f$ uniformly on D then $f_n \to f$ pointwise on D. Uniform convergence is stronger in that given $\varepsilon > 0$, there must exist an N which does the necessary job for all x in D simultaneously, while for pointwise convergence, given $\varepsilon > 0$ we may use a different N_x for each x in D. Thus uniform convergence is global in D.

Here is a graphical interpretation of uniform convergence. Suppose that $D = [0, 1]$, and let us draw the graph of f. Uniform convergence of (f_n) to f on $[0, 1]$ means that if we draw a ribbon of vertical width 2ε centred on the graph of f, then no matter how

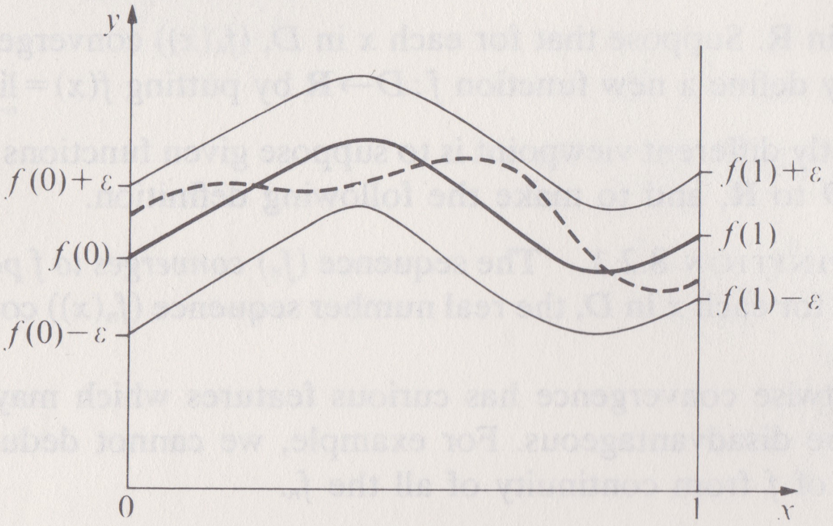


FIG. 8.1

small $\varepsilon > 0$ is, there is some stage in the sequence beyond which the graph of every f_n lies inside this ribbon. (See Fig. 8.1.)

REMARKS 8.2.4(a). Suppose that a sequence of real numbers (a_n) converges to a. For any $D \subset \mathbf{R}$, let $f_n : D \to \mathbf{R}$, $f : D \to \mathbf{R}$ be the constant functions, $f_n(x) = a_n$, $f(x) = a$, for all x in D. Then it is easy to see that (f_n) converges to f uniformly on D.

(b) If $f_n \to f$ uniformly on D, and if $D' \subset D$, then clearly $f_n \mid D' \to f \mid D'$ uniformly on D'.

The first example is a negative one: in Example 8.2.2 convergence is not uniform on $[0, 1]$. To see this, we first put the negation of Definition 8.2.3 into the following usable form: there exists some $\varepsilon_0 > 0$ such that for any N there exists an $n \geqslant N$ and an x in D with $|f_n(x) - f(x)| \geqslant \varepsilon_0$. For Example 8.2.2, we may take $\varepsilon_0 = \frac{1}{2}$. The idea now is to show that no matter how large N is, we can choose an x close to (but not equal to) 1 such that $x^N \geqslant \frac{1}{2}$. Since then $|f_N(x) - f(x)| = |x^N - 0| = x^N \geqslant \frac{1}{2}$, this will show that the convergence is not uniform on $[0, 1]$. If $1 - \delta < x < 1$ (with $\delta < 1$) then $x^N > (1-\delta)^N$, and $(1-\delta)^N \geqslant \frac{1}{2}$ provided $\delta \leqslant 1 - 2^{-1/N}$. Hence we can choose an x as required.

WARNING. The same proof shows that convergence of (x^n) to the zero function is not uniform on $[0, 1)$. It is a tempting fallacy to think that by dropping one 'bad' point, the convergence will become uniform.

On the other hand, if we take $D = [0, a]$ for some a in $(0, 1)$, then it is true that $x^n \to 0$ uniformly on D. For given $\varepsilon > 0$, we may choose N large enough so that $a^N < \varepsilon$, and then for any $n \geqslant N$ and any x in $[0, a]$, $|x^n - 0| \leqslant a^n \leqslant a^N < \varepsilon$.

This successful proof of uniform convergence may be explained as follows. For each fixed n, we were able to locate an upper bound a^n for $|f_n(x) - f(x)|$ on $[0, a]$, and then we used the fact that $a^n \to 0$ as $n \to \infty$. This suggests a criterion for uniform convergence, which is really just a translation of the definition.

PROPOSITION 8.2.5. *Given a sequence* (f_n) *which converges pointwise to* f *on* D, *let* $M_n = \sup_{x \in D} |f_n(x) - f(x)|$ *for each* n *such that the right-hand side exists. Then* $f_n \to f$ *uniformly on* D *iff* $M_n \to 0$ *as* $n \to \infty$.

(Note that '$M_n \to 0$ as $n \to \infty$' includes the information that M_n exists for all sufficiently large n.)

Proof. Suppose that $M_n \to 0$ as $n \to \infty$. Given $\varepsilon > 0$, there exists N such that $0 \leqslant M_n < \varepsilon$ for all $n \geqslant N$. Since, by definition of M_n, $|f_n(x) - f(x)| \leqslant M_n$ for all x in D, we get $|f_n(x) - f(x)| < \varepsilon$ for all $n \geqslant N$ and all x in D.

Conversely, suppose that $f_n \to f$ uniformly on D. Then given $\varepsilon > 0$, there exists N such that $|f_n(x) - f(x)| < \frac{1}{2}\varepsilon$ for all $n \geqslant N$ and all x in D. Thus for $n \geqslant N$, the set $S_n = \{|f_n(x) - f(x)| : x \in D\}$ is bounded above by $\frac{1}{2}\varepsilon$, so $M_n = \sup S_n \leqslant \frac{1}{2}\varepsilon < \varepsilon$, as required.

This criterion has two uses, one practical and one theoretical. On the practical side, along with elementary calculus it often allows us to determine whether particular sequences are uniformly convergent. We give negative and positive examples of this.

EXAMPLE 8.2.6. Let $f_n(x) = n^2 x(1-x)^n$ for x in $[0, 1]$. Then $f_n \to 0$ (the zero function) pointwise on $[0, 1]$: for if $x = 0$ or 1, $f_n(x) = 0$ for all n; and if $x \in (0, 1)$, then $0 < f_n(x) < n^2(1-x)^n$, so $f_n(x) \to 0$ as $n \to \infty$ by Exercise 1.5.9. Let us test for uniform convergence. To find

$$M_n = \sup_{x \in [0,1]} |f_n(x) - f(x)| = \sup_{x \in [0,1]} f_n(x),$$

we use calculus. Since

$$f_n'(x) = n^2(1-x)^{n-1}\{1 - (n+1)x\},$$

$f_n'(x)=0$ when $x=1$, $1/(n+1)$. Moreover,

$$f_n'(x)>0 \qquad \text{for} \qquad 0<x<1/(n+1),$$
$$f_n'(x)<0 \qquad \text{for} \qquad 1/(n+1)<x<1,$$

so M_n is attained at $x=1/(n+1)$. Hence

$$M_n=\frac{n^{n+2}}{(n+1)^{n+1}}\to\infty \qquad \text{as} \qquad n\to\infty,$$

and convergence is not uniform on $[0, 1]$.

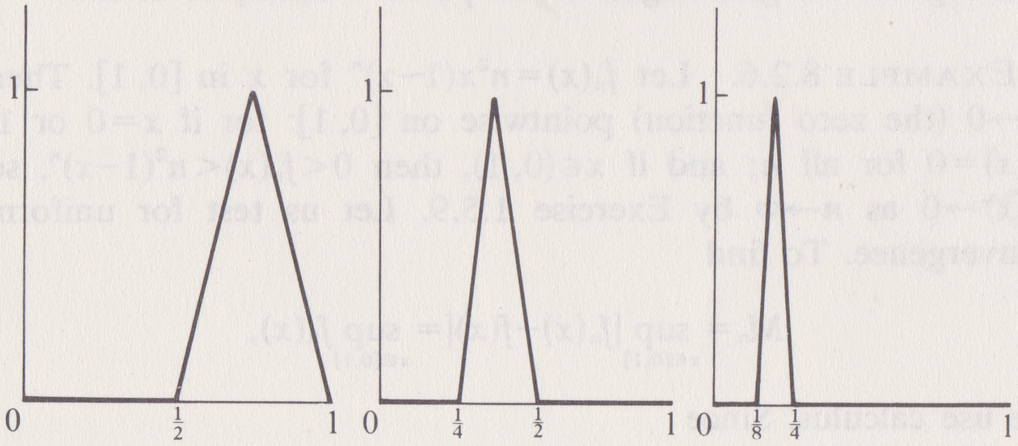
EXAMPLE 8.2.7. Let $f_n(x)=xe^{-nx^2}$, $x\in[0, 1]$. Again we may show that $f_n\to 0$ pointwise on $[0, 1]$. As in Example 8.2.6 we may show that M_n is attained at $x=(2n)^{-\frac{1}{2}}$, and hence that $M_n=(2en)^{-\frac{1}{2}}\to 0$ as $n\to\infty$. Thus convergence is uniform on $[0, 1]$.

We give a final example of non-uniform convergence, in geometric language.

EXAMPLE 8.2.8. (The moving bump.) Let (f_n) be the sequence of functions on $[0, 1]$ such that the graphs of f_1, f_2, f_3 are as shown in Fig. 8.2, and the graph of f_n takes the form suggested by this. (It is easy, but not illuminating, to give a formula for f_n.) For any fixed x in $[0, 1]$, $f_n(x)\to 0$ as $n\to\infty$: for $f_n(0)=0$ for all n, and if $x>0$, then for all sufficiently large n the bump has passed entirely to the left of x, i.e. $f_n(x)=0$, and again $f_n(x)\to 0$ as $n\to\infty$. However,

$$M_n=\sup_{x\in[0,1]}|f_n(x)-0|=1$$

for all n, and the convergence is not uniform. We may see this in



(a) (b) (c)

FIG. 8.2(a)(b)(c)

terms of the 2ε-ribbon picture by taking $\varepsilon<1$; then for any n, the graph of f_n strays outside the ribbon near $x=0$. (The reader may visualize such a ribbon on Fig. 8.2.)

The theoretical use of Proposition 8.2.5 is to re-interpret uniform convergence as (ordinary) convergence in a suitable function space. Let $\mathcal{B}(D, \mathbf{R})$ be the metric space of all bounded real-valued functions on D with the sup metric,

$$d_\infty(f, g)=\sup_{x\in D}|f(x)-g(x)|.$$

Then convergence of a sequence (f_n) in $\mathcal{B}(D, \mathbf{R})$ to a point f means (Definition 4.1.1) that $d_\infty(f_n, f)\to 0$ as $n\to\infty$. But by Proposition 8.2.5 this means precisely that $f_n\to f$ uniformly on D. Convergence in $\mathcal{B}(D, \mathbf{R})$ does not quite cover all the situations allowed in Definition 8.2.3. For example it makes sense (and is true) to say that the sequence $(x^{-1}+n^{-1})$ converges uniformly to x^{-1} on $(0, 1)$, although every function involved is unbounded on $(0, 1)$. In fact uniform convergence for unbounded functions may also be interpreted as convergence in a suitable function space (see for example Choquet (1966) p. 60 ff) but we avoid this technicality by restricting to bounded functions for our function spaces.

8.3 Cauchy's criterion

Recall that in § 1.2 Cauchy sequences were important.

DEFINITION 8.3.1. A sequence (f_n) of real-valued functions on $D\subset\mathbf{R}$ is *uniformly Cauchy on D* if given $\varepsilon>0$, there exists N such that $|f_n(x)-f_m(x)|<\varepsilon$ for all m, $n\geq N$ and all x in D.

As usual, the uniform feature is that for a given $\varepsilon>0$, the same N has to work for all x in D; the condition is global in D. If all the functions concerned are bounded, then as for convergence, a sequence (f_n) is uniformly Cauchy on D iff it is a Cauchy sequence in $\mathcal{B}(D, \mathbf{R})$, in the sense of Definition 7.2.3.

THEOREM 8.3.2. (Cauchy's criterion for uniform convergence.) *Let (f_n) be a sequence of real-valued functions defined on $D\subset\mathbf{R}$. Then (f_n) converges uniformly on D iff (f_n) is uniformly Cauchy on D.*

Proof. The 'only if' part is similar to the corresponding part of the proof of Theorem 1.2.9.

Suppose that (f_n) is uniformly Cauchy on D. For any fixed x in D, $(f_n(x))$ is then a Cauchy sequence in $\mathbf{R}$, so by Theorem 1.2.9 it converges to some real number, say $f(x)$. We shall prove that $f_n \to f$ uniformly on D.

Given $\varepsilon > 0$, choose N so that $|f_n(x) - f_m(x)| < \varepsilon$ for all m, $n \geqslant N$ and all x in D. We give two ways of completing the proof.

(a) For any x in D, by convergence of $(f_n(x))$ to $f(x)$ there exists an integer $m(x)$ such that $|f_{m(x)}(x) - f(x)| < \varepsilon$, and moreover we may choose $m(x) \geqslant N$. Hence for any $n \geqslant N$,

$$|f_n(x) - f(x)| \leqslant |f_n(x) - f_{m(x)}(x)| + |f_{m(x)}(x) - f(x)| < 2\varepsilon.$$

(Note that N is independent of x although $m(x)$ is not.)

(b) Since $|f_n(x) - f_m(x)| < \varepsilon$ for all m, $n \geqslant N$ and all x in D, we may keep n fixed and let $m \to \infty$ to get $|f_n(x) - f(x)| \leqslant \varepsilon$ for all $n \geqslant N$ and all x in D.

Either way, (f_n) converges to f uniformly on D.

REMARKS 8.3.3(a). As for sequences in $\mathbf{R}$, the advantage of Cauchy's criterion is that the limit function need not be known in advance in order to prove uniform convergence. This is particularly useful when we are trying to define a function as the limit of a sequence of functions.

(b) When all the functions involved are bounded, Cauchy's criterion says that the metric space $\mathscr{B}(D, \mathbf{R})$ is like $\mathbf{R}$ in that any Cauchy sequence in it converges. We shall study such spaces further in Chapter 9.

(c) As for sequences in $\mathbf{R}$, one may translate the above results into the language of series, and develop tests for uniform convergence of series of functions. We refer to Ferrar (1938) for this.

8.4 Uniform limits of sequences

THEOREM 8.4.1. *If $f_n : (a, b) \to \mathbf{R}$ is continuous at c in (a, b) for each n and if $f_n \to f$ uniformly on (a, b) then f is continuous at c.*

Proof. This is a 3ε-argument. Given $\varepsilon > 0$, by uniform convergence there exists N such that $|f_n(x) - f(x)| < \varepsilon$ for all $n \geqslant N$ and all x in (a, b). Now f_N is continuous at c, so there exists $\delta > 0$ such that $|f_N(x) - f_N(c)| < \varepsilon$ for all x in $B_\delta(c) \cap (a, b)$. Hence for any such x,

$$|f(x) - f(c)| \leqslant |f(x) - f_N(x)| + |f_N(x) - f_N(c)| + |f_N(c) - f(c)| < 3\varepsilon,$$

which shows that f is continuous at c.

REMARKS 8.4.2(a). Theorem 8.4.1 illustrates the remark in § 8.1 that uniform convergence can validate interchanges of limiting processes: continuity of f at c means that $\lim_{x\to c} f(x)=f(c)$, i.e.

$$\lim_{x\to c}\left(\lim_{n\to\infty} f_n(x)\right)=\lim_{n\to\infty} f_n(c)=\lim_{n\to\infty}\left(\lim_{x\to c} f_n(x)\right),$$

where the second equality follows from continuity of f_n at c.

(b) In terms of function spaces, Theorem 8.4.1 says that in $\mathscr{B}=\mathscr{B}((a, b), \mathbf{R})$ the subspace $\mathscr{C}_c$ of those functions which are continuous at c is closed in $\mathscr{B}$. For if $f\in Cl(\mathscr{C}_c)$ then by Lemma 7.2.2 there exists a sequence (f_n) in $\mathscr{C}_c$ converging to f. Hence by Theorem 8.4.1, $f\in\mathscr{C}_c$. Thus $Cl(\mathscr{C}_c)=\mathscr{C}_c$, so $\mathscr{C}_c$ is closed in $\mathscr{B}$.

We have now carried the theory of uniform convergence far enough for our purposes. For applications to term-by-term differentiation and integration, etc., we refer to Apostol (1957).

*8.5 Generalizations

The definition of uniform convergence makes sense for maps of any set D to any metric space M. In particular, uniform convergence for sequences of bounded maps from D to M may be interpreted as convergence in the metric space $\mathscr{B}(D, M)$ of Example 2.2.17. In Chapter 9 we shall comment on the analogue of Cauchy's criterion in that case. If D is a topological space and $c\in D$, we may consider the subspace $\mathscr{C}_c\subset\mathscr{B}(D, M)$ of those maps continuous at c. The proof of Theorem 8.4.1 readily extends to show that $\mathscr{C}_c$ is closed in $\mathscr{B}(D, M)$. (We need only replace the $B_\delta(c)$ used in that proof by a set U containing c and open in D, such that $f_N(U)\subset B_\varepsilon(f_N(c))$.*

Exercises 8.6

1. Suppose that $D\subset\mathbf{R}$ is of the form $D=\bigcup_{i=1}^{k} D_i$ and that a sequence (f_n) of functions $f_n: D\to\mathbf{R}$ is uniformly convergent on D_i for each $i=1, 2, \ldots, k$. Show that (f_n) is uniformly convergent on D.
2. Show that if $f_n\to f$ uniformly on D and if each f_n is bounded on D then
   (a) f is bounded on D,

(b) there is a uniform bound for the f_n, i.e. there exists K in $\mathbf{R}$ such that $|f_n(x)| \leqslant K$ for all x in D and n in $\mathbf{N}$.

Interpret (b) as a statement about $\mathscr{B}(D, \mathbf{R})$.

3. Which of the following sequences (f_n) converge uniformly on $[0, 1]$?
   (i) $f_n(x) = x/(1+nx)$,
   (ii) $f_n(x) = nxe^{-nx^2}$,
   (iii) $f_n(x) = n^{\frac{1}{2}}x(1-x)^n$,
   (iv) $f_n(x) = nx(1-x^2)^{n^2}$,
   (v) $f_n(x) = x^n/(1+x^n)$,
   (vi) $f_n(x) = n^{-x}x^n \cos nx$.
4. Given that (f_n), (g_n) converge to f, g uniformly on D, prove that $(\lambda f_n + \mu g_n)$ converges to $\lambda f + \mu g$ uniformly on D, for any λ, μ in $\mathbf{R}$.
5. Given that (f_n), (g_n) converge to f, g uniformly on D, and that for each n the functions f_n and g_n are bounded, prove that $f_n g_n$ converges uniformly to fg. (Hint: use Exercise 2(b).)

6*. Let

$$f_n(x) = (1+n^{-1})x,$$

$$g_n(x) = \begin{cases} n^{-1} & \text{if} \quad x = 0 \quad \text{or} \quad x \notin \mathbf{Q}, \\ q + n^{-1} & \text{if} \quad x = p/q \text{ where } p \in \mathbf{Z},\ q \in \mathbf{N} \text{ and } p,\ q \\ & \qquad\qquad \text{have highest common factor } 1, \end{cases}$$

$$h_n(x) = f_n(x)g_n(x).$$

Show that (f_n), (g_n) converge uniformly on any interval $[a, b]$, but that (h_n) does not converge uniformly on any interval containing more than one point.

7. Construct a sequence of functions $f_n : \mathbf{R} \to \mathbf{R}$ none of which is continuous at $x = 0$, but such that (f_n) converges uniformly on $\mathbf{R}$ to a continuous function.
8. Suppose that (f_n) converges uniformly on D to a function f, and that f_n is uniformly continuous on D for each n. Prove that f is uniformly continuous on D.

9*. (Dini's theorem.) Given that (f_n) is a sequence of continuous real-valued functions on a compact topological space T such that $f_n(x) \geqslant f_{n+1}(x)$ for all n and for all x in T, and that (f_n) converges pointwise to a continuous function f on T, prove that (f_n) converges uniformly on T.

9

Complete metric spaces

9.1 Introduction

We have already seen the usefulness of the completeness axiom for **R**. In more general contexts, it is often desirable to have an analogous completeness property. Axiom 1.1.4 is convenient for studying continuity of real-valued functions of a real variable, but it is too closely associated with the order properties of **R** for generalizing to metric spaces. However Cauchy's criterion, which we saw in § 1.2 is closely related to the completeness axiom, makes sense in any metric space; if it is true, the space is called complete. In this chapter we study complete metric spaces and give some applications. Completeness will be used again in Chapter 10.

9.2 Definition and examples

Definition 9.2.1. A metric space M is *complete* if every Cauchy sequence in M converges (to a point in M).

Examples 9.2.2(a). **R** is complete by Theorem 1.2.9.

(b) **Q** is not complete: any sequence in **Q** which converges in **R** to $\sqrt{2}$ is a Cauchy sequence in **Q** which does not converge to any point in **Q**.

(c) $(0, 1) \subset \mathbf{R}$ is not complete: $(1/n)$ is a Cauchy sequence in $(0, 1)$ which does not converge to any point in $(0, 1)$.

Examples 9.2.2.(a) and (c) show that completeness is not a topological invariant. However (see Exercise 9.6.4) it is invariant under uniform equivalence (hence in particular under Lipschitz equivalence) of metrics. As Examples 9.2.2(b) and (c) suggest, any non-closed subspace of a metric space is not complete. This follows from (a) of the next proposition.

Proposition 9.2.3(a). *A complete subspace M_1 of a metric space M_2 is closed in M_2.*

(b) *A closed subspace M_1 of a complete metric space M_2 is complete.*

Proof. (a) Suppose that M_1 is complete and that $x \in Cl(M_1)$. By Lemma 7.2.2, there exists a sequence (x_n) in M_1 converging to x. Since (x_n) is convergent it is Cauchy, so by completeness of M_1, (x_n) converges to a point in M_1. By uniqueness of limits of sequences in metric spaces, this point must be x. Hence $x \in M_1$. This proves that $Cl(M_1) \subset M_1$, so M_1 is closed in M_2.

(b) Suppose that M_2 is complete and that M_1 is closed in M_2. Then $Cl(M_1) = M_1$. Let (x_n) be a Cauchy sequence in M_1. By completeness of M_2, (x_n) converges to a point x in M_2. By Lemma 7.2.2, $x \in Cl(M_1)$. Hence $x \in M_1$, and M_1 is complete.

PROPOSITION 9.2.4. *Any compact metric space is complete.*

Proof. Suppose that M is compact and that (x_n) is a Cauchy sequence in M. By compactness, (x_n) has a subsequence, say $(x_{n(r)})$, converging to a point x in M. Then (x_n) converges to x by the following lemma, which completes the present proof.

LEMMA 9.2.5. *If a Cauchy sequence in a metric space has a subsequence converging to x, then (x_n) converges to x.*

Proof. Let the metric be d, let the subsequence be $(x_{n(r)})$, and let $\varepsilon > 0$. Since (x_n) is Cauchy, there exists N such that $d(x_m, x_n) < \varepsilon$ for all $m, n \geqslant N$. Since $(x_{n(r)})$ converges to x, there exists R such that $d(x_{n(r)}, x) < \varepsilon$ for all $r \geqslant R$. For any $n \geqslant N$, choose $r \geqslant R$ such that $n(r) \geqslant N$, and then

$$d(x_n, x) \leqslant d(x_n, x_{n(r)}) + d(x_{n(r)}, x) < 2\varepsilon.$$

Hence (x_n) converges to x, as required.

Proposition 9.2.4 is not very strong (it is ‘easier’ for a metric space to be complete than to be compact) but it will be of interest again in Chapter 10.

EXAMPLE 9.2.6. $\mathbf{R}^n$ is complete.

Proof. It would be fractionally easier to use the max metric d_∞ than the Euclidean d_2, but for the sake of the next example we use d_2. Suppose that (x^r) is a Cauchy sequence in $\mathbf{R}^n$, where

$$x^r = (x_1^r, x_2^r, \ldots, x_n^r).$$

(Notice that n is fixed in this argument, and that the superscripts are not powers but labels indicating place in the sequence.)

For each $i=1, 2,\ldots, n$, (x_i^r) is a Cauchy sequence in $\mathbf{R}$, since $|x_i^r-x_i^s|\leqslant d_2(x^r, x^s)$ by definition of d_2. Hence (x_i^r) converges, to x_i say. We shall prove that (x^r) converges to $x=(x_1, x_2,\ldots, x_n)$ in $\mathbf{R}^n$. Given $\varepsilon>0$, for each i there exists N_i such that $|x_i^r-x_i|<\varepsilon$ for all $r\geqslant N_i$. Then for all $r\geqslant \text{Max}\{N_1, N_2,\ldots, N_n\}$,

$$d_2(x^r, x)=\left\{\sum_{i=1}^{n}(x_i^r-x_i)^2\right\}^{\frac{1}{2}}<\varepsilon n^{\frac{1}{2}}.$$

Since n is a constant here, (x^r) converges to x in $\mathbf{R}^n$.

REMARK. The same proof shows that any finite product of complete metric spaces is complete. It is also true (Exercise 9.6.7) that if a product of metric spaces is complete then so are the factors.

*EXAMPLE 9.2.7. Hilbert space ℓ^2 is complete.

Proof. Let d_2 denote the metric in ℓ^2. Let (x^s) be a Cauchy sequence in ℓ^2, where

$$x^s=(x_1^s, x_2^s,\ldots, x_n^s,\ldots),$$

the superscripts again being labels, not powers. As in Example 9.2.6, (x_i^s) converges to x_i, say, for each i. We shall prove that $x=(x_1, x_2,\ldots, x_n,\ldots)$ lies in ℓ^2 and that (x^s) converges to x in ℓ^2. Let $\varepsilon>0$. By the Cauchy condition, there exists N such that $d_2(x^r, x^s)<\varepsilon$ for all $r, s\geqslant N$. Hence for all n,

$$\left\{\sum_{i=1}^{n}(x_i^r-x_i^s)^2\right\}^{\frac{1}{2}}\leqslant d_2(x^r, x^s)<\varepsilon \quad \text{for} \quad r, s\geqslant N,$$

where the first inequality follows from the definition of d_2. Keeping s, n fixed and letting $r\to\infty$, we get

$$\left\{\sum_{i=1}^{n}(x_i-x_i^s)^2\right\}^{\frac{1}{2}}\leqslant\varepsilon \qquad \text{for all } s\geqslant N, \text{ and for all } n.$$

Since this holds for all n, it follows that the series of positive terms $\sum_{i=1}^{\infty}(x_i-x_i^s)^2$ converges to some limit not greater than ε^2. Thus for fixed $s\geqslant N$ the real number sequence (y_i) with $y_i=x_i-x_i^s$ lies in ℓ^2, and hence by Example 2.2.16 $x=(x_i)=(y_i+x_i^s)$ is in ℓ^2. Moreover, we saw above that

$$d_2(x, x^s) = \left\{ \sum_{i=1}^{\infty} (x_i - x_i^s)^2 \right\}^{\frac{1}{2}} \leqslant \varepsilon \qquad \text{for all } s \geqslant N,$$

hence (x^s) converges to x in ℓ^2.*

EXAMPLE 9.2.8. The space $\mathscr{B}(D, \mathbf{R})$ of bounded real-valued functions on $D \subset \mathbf{R}$ with the sup metric is complete. This has already been noted in § 8.3.

*EXAMPLE 9.2.9. More generally, as in § 8.5 consider the analogous space $\mathscr{B}(D, M)$ for any set D and any metric space M.

PROPOSITION 9.2.10. *$\mathscr{B}(D, M)$ is complete iff M is complete.*

Proof. A close scrutiny of the proof of Theorem 8.3.2 reveals that it uses nothing about D, and only the completeness of $\mathbf{R}$, so the same proof shows that $\mathscr{B}(D, M)$ is complete provided M is complete. Conversely if M is not complete, let (x_n) be a non-convergent Cauchy sequence in M. The sequence (f_n) of constant functions given by $f_n(y) = x_n$ for all y in D is then a non-convergent Cauchy sequence in $\mathscr{B}(D, M)$ since (f_n) does not converge even pointwise on D.*

EXAMPLES 9.2.11. Let $\mathscr{B}(D, \mathbf{R})$ be as in Example 9.2.8, and let $\mathscr{C}_c$ be the subspace of those functions which are continuous at some given point c in D. By § 8.4, $\mathscr{C}_c$ is a closed subspace of the complete space $\mathscr{B}(D, \mathbf{R})$ and hence is complete. Let $\mathscr{C}(D, \mathbf{R})$ be the subspace of all continuous bounded functions in $\mathscr{B}(D, \mathbf{R})$. Then

$$\mathscr{C}(D, \mathbf{R}) = \bigcap_{c \in D} \mathscr{C}_c.$$

Hence $\mathscr{C}(D, \mathbf{R})$ is also closed in $\mathscr{B}(D, \mathbf{R})$ and is complete.

More generally, if T is any topological space and M is any complete metric space then the space $\mathscr{B}(T, M)$ of all continuous bounded maps from T to M is complete by the same argument.

EXAMPLE 9.2.12. The space M of continuous real-valued functions on $[0, 1]$ with the L^1 metric d_1 (Example 2.2.9) is not complete. To see this, let $f_n:[0,1]\to\mathbf{R}$ be given by

$$f_n(x)=\begin{cases}0, & x\in\left[0,\dfrac{1}{2}-\dfrac{1}{n}\right],\\ 1, & x\in[\tfrac{1}{2},1],\\ n\left(x+\dfrac{1}{n}-\dfrac{1}{2}\right), & x\in\left[\dfrac{1}{2}-\dfrac{1}{n},\dfrac{1}{2}\right],\end{cases}$$

for any $n\geqslant 2$. Then (f_n) is a Cauchy sequence in M, for given $\varepsilon>0$, choose $N>1/\varepsilon$ and it follows that

$$d_1(f_n,f_m)=\int_0^1|f_n-f_m|<\varepsilon\quad\text{for all } m>n\geqslant N,$$

since $f_n(x)=f_m(x)$ except in $\left[\dfrac{1}{2}-\dfrac{1}{n},\dfrac{1}{2}\right]$ and $|f_n(x)-f_m(x)|\leqslant 1$ for x in $\left[\dfrac{1}{2}-\dfrac{1}{n},\dfrac{1}{2}\right]$ (see Fig. 9.1). We shall prove that (f_n) does not converge in M. First here are two fallacious proofs.

(a) For any x in $[0,\tfrac{1}{2})$, $f_n(x)\to 0$ as $n\to\infty$, and for x in $[\tfrac{1}{2},1]$, $f_n(x)\to 1$ as $n\to\infty$. Hence the limit function f of (f_n) is given by

$$f(x)=\begin{cases}0, & x\in[0,\tfrac{1}{2}),\\ 1, & x\in[\tfrac{1}{2},1].\end{cases}$$

This is not continuous on $[0,1]$, hence f is not in M, so (f_n) does not converge in M.

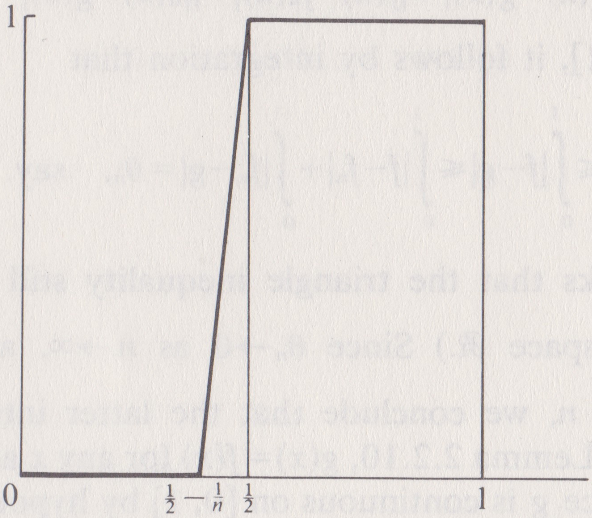


FIG. 9.1

What this proves is that $f_n \to f$ pointwise on $[0, 1]$. It is still conceivable that there is some other, continuous, function to which (f_n) converges in the L^1 metric. To clarify this fallacy, suppose we let $g_n = f_n \mid [0, \frac{1}{2}]$, and consider the sequence (g_n) in the space M' similar to M but with $[0, 1]$ replaced by $[0, \frac{1}{2}]$. Then the argument in (a) would say that (g_n) converges to the discontinuous function $f \mid [0, \frac{1}{2}]$, hence (g_n) does not converge in M'. But it fact (g_n) does converge in M', to the zero function g, for $\int_0^{\frac{1}{2}} |g_n - g| = 1/2n \to 0$ as $n \to \infty$.

(b) Let f be as in (a). Then $\int_0^1 |f_n - f| = 1/2n \to 0$ as $n \to \infty$. Hence (f_n) converges to f in the L^1 metric, and since limits of sequences in a metric space are unique, (f_n) cannot converge to any continuous function in the L^1 metric.

This is closer; at least it uses the correct metric. But it will still not do, as it stands. For f belongs to, say, the set $\mathcal{R}$ of Riemann-integrable functions on $[0, 1]$. (Note that $M \subset \mathcal{R}$.) Now the formula $d_1(g, h) = \int_0^1 |g - h|$ does not define a metric in this larger set $\mathcal{R}$, but only a so-called pseudo-metric: $\int_0^1 |g - h| = 0$ does not imply $g = h$ in $\mathcal{R}$ (g and h might for example differ at a finite set of points).

(c) However, we can extend (b) to a valid proof. Let f be as before; thus $\int_0^1 |f_n - f| \to 0$ as $n \to \infty$. Suppose that (f_n) converges to a continuous function g in the L^1 metric, i.e. $\int_0^1 |f_n - g| \to 0$ as $n \to \infty$. Since

$$|f(x) - g(x)| \leqslant |f(x) - f_n(x)| + |f_n(x) - g(x)|$$

for all x in $[0, 1]$, it follows by integration that

$$0 \leqslant \int_0^1 |f - g| \leqslant \int_0^1 |f - f_n| + \int_0^1 |f_n - g| = \theta_n, \quad \text{say.}$$

(This just checks that the triangle inequality still holds in the pseudo-metric space $\mathcal{R}$.) Since $\theta_n \to 0$ as $n \to \infty$, and $\int_0^1 |f - g|$ is independent of n, we conclude that the latter integral is zero. By the proof of Lemma 2.2.10, $g(x) = f(x)$ for any x at which $f - g$ is continuous. Since g is continuous on $[0, 1]$ by hypothesis and f is continuous except at $x = \frac{1}{2}$, this means that $g(x) = f(x)$ for $x \neq \frac{1}{2}$, which is clearly impossible for a continuous g. Hence M is not complete.

In the appendix we show how any incomplete metric space may be 'completed' in an essentially unique way. This is closely related to one way of constructing **R** from **Q**. The process of completing Example 9.2.12 contains some (though not all) of the idea of the Lebesgue integral.

9.3 Fixed point theorems

One of the most attractive theorems about metric spaces is Banach's fixed point theorem. On the theoretical side, it unifies the proofs of several theorems on the existence of solutions of algebraic, differential, integral, and functional equations. The theorem, when it works, tells us that the solution is unique as well. It employs a constructive method, and is of interest also in numerical analysis. In this section we describe a special case which may or may not be found helpful preparation for the general theorem in § 9.4. (Banach's theorem may be clearer to some readers when it is stated in general.)

DEFINITION 9.3.1. Given any map $f:S\to S$ of a set S, a *fixed point of* f is a point p in S such that $f(p)=p$.

Conscientious readers will already have proved that any continuous $f:[a,b]\to[a,b]$ has a fixed point (Exercise 6.6.10). A special case of Banach's theorem proves, under stronger hypotheses, that this fixed point is unique. We need some terminology. Let $D\subset\mathbf{R}$.

DEFINITION 9.3.2. A function $f:D\to\mathbf{R}$ *satisfies a Lipschitz condition of order* α (*with constant* K) *on* D if

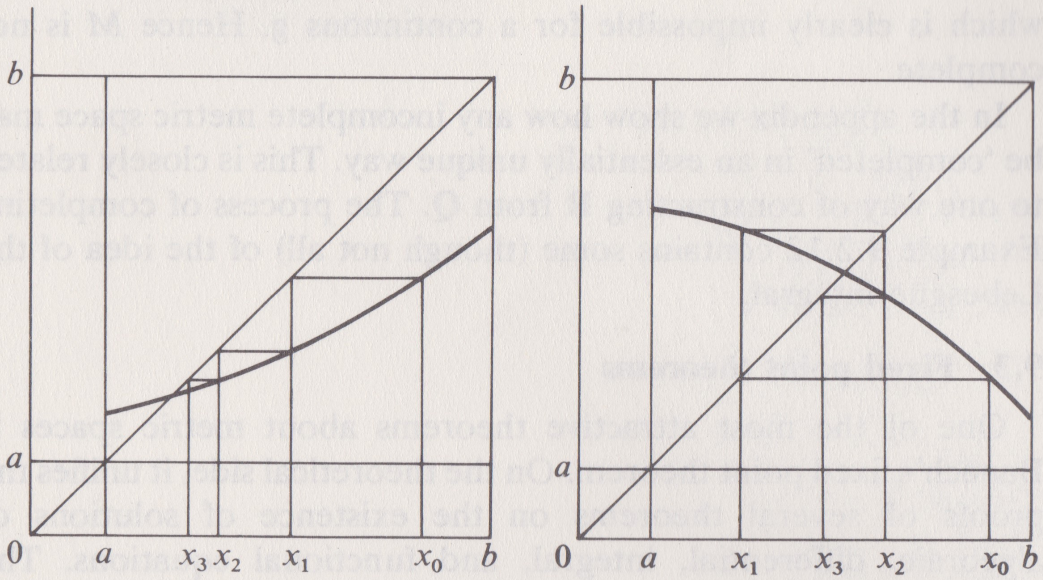
$$|f(x)-f(y)|\leq K|x-y|^{\alpha} \qquad \text{for all } x, y \text{ in } D.$$

Exercise 9.6.11 asks the reader to relate this to more familiar properties, as follows.

PROPOSITION 9.3.3(a). *If f satisfies a Lipschitz condition of order $\alpha>0$ on D, then f is uniformly continuous on D.*

(b) *If f satisfies a Lipschitz condition of order $\alpha>1$ on $[a,b]$, then f is constant on $[a,b]$.*

(c) *If f is continuous on $[a,b]$ and differentiable on (a,b) with $|f'(x)|\leq K$ for all x in (a,b) then f satisfies a Lipschitz condition of order 1 with constant K on $[a,b]$.*



(a) (b)

FIG. 9.2(a)(b)

THEOREM 9.3.4. *If $f:[a, b]\rightarrow[a, b]$ satisfies a Lipschitz condition of order 1 with constant $K<1$ on $[a, b]$, then f has a unique fixed point p in $[a, b]$. If x_0 is any point in $[a, b]$ and $x_n=f(x_{n-1})$ for all $n\geqslant 1$, then (x_n) converges to p.*

Since this is a special case of Banach's theorem in § 9.4, we shall not prove it now. Figure 9.2 illustrates two possible cases.

9.4 The contraction mapping theorem

Lipschitz conditions generalize easily to metric spaces. The kind of condition we are concerned with has an appealing geometric interpretation.

DEFINITION 9.4.1. A map $f:M\rightarrow M$ of a metric space M with metric d is a *contraction* if there exists a constant $K<1$ such that $d(f(x), f(y))\leqslant Kd(x, y)$ for all x, y in M.

LEMMA 9.4.2. *Any contraction is uniformly continuous.*

Proof. With the above notation, let $\varepsilon>0$ and take $\delta=\varepsilon/K$. For any x, y in M satisfying $d(x, y)<\delta$, we have

$$d(f(x), f(y))\leqslant Kd(x, y)<\varepsilon,$$

as required.

THEOREM 9.4.3(Banach). *If $f:M\rightarrow M$ is a contraction of a complete metric space M, then f has a unique fixed point p in M.*

Proof. *Existence.* Choose x_0 arbitrarily in M, and inductively let $x_n = f(x_{n-1})$ for $n \geqslant 1$. We shall prove that (x_n) is a Cauchy sequence.

First we note that $d(x_r, x_{r-1}) \leqslant K^{r-1} d(x_1, x_0)$ for all $r \geqslant 1$. This follows easily by induction from the contraction condition. Now for $m > n$, by repeated use of the triangle inequality,

$$d(x_m, x_n) \leqslant d(x_m, x_{m-1}) + d(x_{m-1}, x_{m-2}) + \ldots + d(x_{n+1}, x_n).$$

Hence

$$\begin{aligned} d(x_m, x_n) &\leqslant (K^{m-1} + K^{m-2} + \ldots + K^n) d(x_1, x_0) \\ &= K^n (K^{m-n-1} + K^{m-n-2} + \ldots + 1) d(x_1, x_0) \\ &= \frac{K^n (1 - K^{m-n})}{1-K} d(x_1, x_0) \\ &< \frac{K^n}{1-K} d(x_1, x_0). \end{aligned}$$

Since $0 \leqslant K < 1$, $K^n \to 0$ as $n \to \infty$. Hence (x_n) is a Cauchy sequence. Since M is complete, (x_n) converges to some p in M. By continuity of f at p (Lemma 9.4.2), $f(x_n) \to f(p)$ as $n \to \infty$. But $f(x_n) = x_{n+1} \to p$ as $n \to \infty$. It follows that $f(p) = p$.

Uniqueness. If $f(p) = p$, $f(q) = q$, then

$$d(p, q) = d(f(p), f(q)) \leqslant K d(p, q).$$

Since $K < 1$, this is a contradiction unless $d(p, q) = 0$. Hence $q = p$, and the fixed point of f is unique.

From the practical viewpoint, the next result is also of interest.

PROPOSITION 9.4.4. *With the notation and hypotheses of Theorem* 9.4.3,

$$d(x, p) \leqslant \frac{1}{1-K} d(f(x), x)$$

for any x in M.

This is proved by putting $x_0 = x$ in the proof of the theorem. We omit the details. Proposition 9.4.4 may be thought of as an error estimate: without yet knowing the correct p, we get an estimate of how far wrong any given x is. For example if for some choice of x_0 we put $x = x_n$ where x_n is as in the proof of the theorem, then

$$d(x_n, p) \leqslant \frac{1}{1-K} d(f(x_n), x_n) \leqslant \frac{K^n}{1-K} d(x_1, x_0).$$

REMARKS 9.4.5(a). There are many variations on the contraction mapping theme. For example it is enough to assume, instead of the contraction condition, that some iterate $f^{(r)}$ of f is a contraction (see Exercise 9.6.15). However, as Exercises 9.6.12, 9.6.13 show, we cannot simply delete completeness or the uniform factor K.

(b) The statement that f is a contraction depends heavily on the particular metric used. It is possible that d, d' are uniformly equivalent metrics (so that M is d-complete iff M is d'-complete) and that f is a d-contraction but not a d'-contraction. (A trivial example of this occurs in Exercise 9.6.17.) This gives scope for ingenuity in the choice of metric.

One application of the contraction mapping theorem has been given in Theorem 9.3.4. We refer the reader to Kolmogorov and Fomin (1957) for applications to systems of equations, differential and integral equations, and to Dieudonné (1960) for an application to implicit (and inverse) function theorems. The theorem unifies various methods of successive approximations which were originally used to prove such results. Kolmogorov and Fomin give a version of Picard's theorem on the local existence and uniqueness of a solution to the initial value problem

$$\frac{dy}{dx} = f(x, y); \qquad y(x_0) = y_0,$$

under suitable conditions on f. 'Local' here means in a neighbourhood of x_0. The contraction mapping theorem applied using the sup metric on a suitable function space gives the result in a neighbourhood smaller than the one in Picard's original theorem. We refer to Pitts (1972) for an illustration of Remark 9.4.5(b); a variant of the sup metric is used to recover Picard's original range of validity.

To indicate how these applications go, we give one example. Let us consider the (possibly unfamiliar) equation

$$\psi(x) = \lambda \int_a^b K(x, y)\psi(y)\, dy + \theta(x),$$

where K and θ are known, and the problem is to find ψ. This is called an integral equation (Fredholm, of the second kind). Suppose that $K : [a, b] \times [a, b] \to \mathbf{R}$, $\theta : [a, b] \to \mathbf{R}$ are continuous and that $\lambda \in \mathbf{R}$.

We shall prove that provided $|\lambda|$ is sufficiently small, there is a unique continuous $\psi : [a, b] \to \mathbf{R}$ satisfying the integral equation.

Let $M=\mathscr{C}[a, b]$ be the metric space in Example 2.2.8, with the sup metric d_∞. By Example 9.2.11 this is complete. Define $f:M\to M$ as follows: for any ψ in M, Let $f(\psi):[a, b]\to\mathbf{R}$ be given by

$$f(\psi)(x)=\lambda\int_a^b K(x, y)\psi(y)\,dy+\theta(x),$$

for all x in $[a, b]$. It follows easily from integration theory that $f(\psi)$ is continuous on $[a, b]$, so $f(\psi)\in M$. We shall prove that for $|\lambda|$ sufficiently small, f is a contraction.

Since $[a, b]\times[a, b]$ is compact and K is continuous, we may choose Δ such that $|K(x, y)|\leqslant\Delta$ for all x, y in $[a, b]$. For any ψ_1, ψ_2 in M,

$$\begin{aligned} d_\infty(f(\psi_1), f(\psi_2)) &= \sup_{x\in[a,b]} |f(\psi_1)(x)-f(\psi_2)(x)| \\ &= \sup_{x\in[a,b]} \left|\lambda\int_a^b K(x, y)(\psi_1(y)-\psi_2(y))\,dy\right| \\ &\leqslant|\lambda|\,(b-a)\Delta \sup_{y\in[a,b]} |\psi_1(y)-\psi_2(y)|, \end{aligned}$$

where the last inequality follows from integration theory. Since

$$\sup_{y\in[a,b]} |\psi_1(y)-\psi_2(y)|=d_\infty(\psi_1, \psi_2),$$

it follows that f is a contraction provided $|\lambda|<\{(b-a)\Delta\}^{-1}$. Hence by the contraction mapping theorem, f has a unique fixed point in M, for such λ. But $f(\psi)=\psi$ means that for all x in $[a, b]$,

$$\lambda\int_a^b K(x, y)\psi(y)\,dy+\theta(x)=\psi(x).$$

Thus for $|\lambda|<\{(b-a)\Delta\}^{-1}$ the equation has a unique continuous solution. (Note that this means 'unique among continuous functions.')

*9.5 Cantor's and Baire's theorems

In this section we treat briefly two important theorems about completeness.

THEOREM 9.5.1 (Cantor). *A metric space M is complete iff for any sequence (V_n) of non-empty closed subsets of M such that $V_n \supset V_{n+1}$ for all n, and such that* diam $V_n \to 0$ *as $n \to \infty$, we have* $\bigcap_{n=1}^{\infty} V_n \neq \varnothing$.

REMARK. If diam $V_n \to 0$ as $n \to \infty$, then $\bigcap_{n=1}^{\infty} V_n$ contains at most one point.

Proof. Suppose that M is not complete, and let (x_n) be a non-convergent Cauchy sequence in M. Let $V_n = \{x_r : r \geqslant n\}$. By Lemmas 7.2.5 and 9.2.5 the set V_n has no limit point in M, so V_n is closed in M. Clearly $V_n \supset V_{n+1}$. Also, given $\varepsilon > 0$, there exists N such that $d(x_m, x_n) < \varepsilon$ for all $m, n \geqslant N$, so diam $V_N \leqslant \varepsilon$, and diam $V_n \to 0$ as $n \to \infty$. Suppose that $x \in V_n$ for all n. Then (x_n) converges to x, since for any n, $d(x, x_n) \leqslant$ diam V_n, so $d(x, x_n) \to 0$ as $n \to \infty$. This contradiction shows that $\bigcap_{n=1}^{\infty} V_n = \varnothing$.

Suppose that M is complete and let V_n be a sequence as in the statement of the theorem. Let $x_n \in V_n$. Then (x_n) is a Cauchy sequence, since given $\varepsilon > 0$, there exists N such that diam $V_N < \varepsilon$, and then for any $m, n \geqslant N$, $x_m, x_n \in V_N$ so $d(x_m, x_n) < \varepsilon$. By completeness of M, (x_n) converges to x, say. Since $x_n \in V_m$ for all $n \geqslant m$, $x \in Cl(V_m) = V_m$ by Lemma 7.2.2. Since this is true for any m, $x \in V_m$ for all m as required.

Exercise 9.6.19 shows that Cantor's theorem is 'sharp'.

In order to state Baire's theorem, recall that a subspace H of a space T is nowhere dense in T if $\mathrm{Int}(Cl(H)) = \varnothing$. Baire's theorem says that a countable collection of nowhere dense subsets is not enough to fill up a complete metric space. Before giving the formal statement (which is stronger) it is convenient to characterize nowhere dense sets in a metric space as follows.

LEMMA 9.5.2. *A subset H of a metric space M is nowhere dense in M iff every non-empty open set in M contains an open ball disjoint from H.*

Proof. Suppose that the latter condition holds, and let $x \in M$, $\varepsilon > 0$. Then $B_\varepsilon(x)$ contains an open ball, $B_\delta(y)$ say, disjoint from H. Thus $y \notin Cl(H)$, so no open ball $B_\varepsilon(x)$ is contained in $Cl(H)$, and $\mathrm{Int}(Cl(H)) = \varnothing$.

Conversely suppose that $\mathrm{Int}(Cl(H))=\varnothing$ and that U is any non-empty open set in M. Let $x\in U$. Then $B_\varepsilon(x)\subset U$ for some $\varepsilon>0$. Since $x\notin \mathrm{Int}(Cl(H))$, there exists y in $(M-Cl(H))\cap B_\varepsilon(x)$. Since this intersection is open, there exists $\delta>0$ such that $B_\delta(y)\subset (M-Cl(H))\bigcap B_\varepsilon(x)$. Then $B_\delta(y)\cap H=\varnothing$ and $B_\delta(y)\subset B_\varepsilon(x)\subset U$, as required.

THEOREM 9.5.3 (Baire). *Let* (H_n) *be a sequence of nowhere dense subsets of a complete metric space* M, *and put* $H=\bigcup_{n=1}^{\infty} H_n$. *Then* $M-H$ *is everywhere dense in* M.

Proof. We have to show that any non-empty open $U\subset M$ contains a point of $M-H$ (since then $Cl(M-H)=M$). First, since H_1 is nowhere dense in M, U contains an open ball $B_{r_1}(x_1)$ disjoint from H_1. We may assume that $r_1<1$. In order to apply Cantor's theorem we want to deal with closed sets, so let V_1 be the 'concentric' closed set $V_1=\{y\in M: d(x_1, y)\leqslant\frac{1}{2}r_1\}$. As noted in Example 3.7.22, V_1 may be strictly larger than $Cl(B_{\frac{1}{2}r_1}(x_1))$, but at least V_1 is a closed set contained in $B_{r_1}(x_1)$ and hence disjoint from H_1.

Now suppose inductively that closed sets $V_1\supset V_2\supset\ldots\supset V_n$ have been chosen so that $V_i\cap H_i=\varnothing$ for $i=1, 2,\ldots, n$, and V_n contains a non-empty set open in M. Then by Lemma 9.5.2, V_n contains some open ball $B_{r_{n+1}}(x_{n+1})$ disjoint from H_{n+1}, and we may assume that $r_{n+1}<1/(n+1)$. Now

$$V_{n+1}=\{y\in M: d(x_{n+1}, y)\leqslant\tfrac{1}{2}r_{n+1}\}$$

is a closed set disjoint from H_{n+1}, with $V_n\supset V_{n+1}$, and V_{n+1} contains the non-empty open set $B_{\frac{1}{2}r_{n+1}}(x_{n+1})$. This defines inductively a nested sequence of non-empty closed sets (V_n) with diam $V_n\to 0$ as $n\to\infty$, so by Cantor's theorem, $\bigcap_{n=1}^{\infty} V_n=\{z\}$ say. Since $z\in V_n$ and $V_n\cap H_n=\varnothing$, $z\in(M-H)\cap U$ as required.

Baire's theorem is a non-constructive existence theorem: it asserts the existence of a point z (indeed, a dense set of points) with the property that z is not in H_n for any n, but in contrast to the contraction mapping theorem, it does not tell us how to find z. For applications of Baire's theorem at a reasonably elementary level, the reader is referred to Pitts (1972) or Boas (1969). There are other equivalent statements of the theorem. If H_n is nowhere dense in M, then $Cl(H_n)$ is nowhere dense in M, since $Cl(Cl(H_n))=Cl(H_n)$. Taking complements in Theorem 9.5.3, we get that the intersection of a sequence of everywhere dense open sets in a complete metric space is everywhere dense. The qualification 'open' is necessary here to ensure that the complement is nowhere dense (see Example 3.7.30).*

Exercises 9.6

1. Show that in any metric space
   (a) the union of two complete subspaces is complete,
   (b) the intersection of any family of complete subspaces is complete (provided it is non-empty).
   (Hint: use Lemma 9.2.5 for (a).)
2. Prove that any discrete metric space in the sense of Example 2.2.2 is complete.
3. Which of the following are complete?
   (i) $\{1/n : n \in \mathbf{N}\} \cup \{0\}$ (ii) $\mathbf{Q} \cap [0, 1]$,
   (iii) $\{(x, y) \in \mathbf{R}^2 : x > 0, y \geqslant 1/x\}$.
4. Suppose that $f : M \to M'$ is a one–one correspondence of metric spaces such that f is uniformly continuous and f^{-1} is continuous.
   (a) Prove that if (x_n) is a Cauchy sequence in M then $(f(x_n))$ is a Cauchy sequence in M'.
   (b) Recall from Exercise 7.3.4 that if (y_n) is a convergent sequence in M' then $(f^{-1}(y_n))$ is a convergent sequence in M.
   (c) Deduce that if M' is complete then M is complete.
   (d) Deduce that completeness is an invariant of uniform equivalence.
5. Give an example of topologically equivalent metrics d, d' on a set A and a sequence (x_n) in A which is Cauchy in $\{A, d\}$ but not in $\{A, d'\}$. (Hint ; see Exercise 5.10.25.)
6. Show that the uniformly continuous image of a complete metric space is not necessarily complete.
7. Prove that if the product $M_1 \times M_2$ of two metric spaces is complete then so are M_1 and M_2. Deduce the analogue for any finite product.
8. Show that the metric space in Example 2.2.11 is not complete.
9. Let M denote the space of bounded real sequences with the

sup metric, and let $C \subset M$ be the set of convergent sequences. Prove that C is closed in M and deduce that C is complete.

10. Let $F \subset \ell^2$ denote the set of (x_n) with $x_n = 0$ for all but a finite number of values of n. Show that F is not complete.

11. Prove Proposition 9.3.3. (Hint: use the mean value theorem for (b) and (c).)

12. Let $f:(0, \frac{1}{4}) \to (0, \frac{1}{4})$ be given by $f(x) = x^2$. Show that f is a contraction without a fixed point.

13. Let $f:[1, \infty) \to [1, \infty)$ be given by $f(x) = x + x^{-1}$. Show that for any distinct x, y, $|f(x) - f(y)| < |x - y|$, that $[1, \infty)$ is complete, but that f has no fixed point.

14. Do Exercise 1.5.14 by showing that $f(x) = (2 + x^{\frac{1}{2}})^{\frac{1}{2}}$ defines a contraction on a suitable closed interval.

15. Suppose that $f: M \to M$ is a map of a complete metric space M, and that for some integer r, the iterated map $f^{(r)} = f \circ f \circ \ldots \circ f$ (r times) is a contraction. Prove that f has a unique fixed point p in M, and* that for any x in M the sequence $(f^{(n)}(x))$ converges to p.

16. Show that the cosine function $\cos : \mathbf{R} \to \mathbf{R}$ is not a contraction, but that $\cos^{(2)}$ (in the notation of Exercise 15) is a contraction. Find an approximate solution of $x = \cos x$.

17. Let $A = \{x, y, z\}$, let d be the discrete metric (Example 2.2.2) on A, and let

$$d'(x, y) = 2, \quad d'(x, z) = 2, \quad d'(y, z) = 1,$$
$$d'(x, x) = d'(y, y) = d'(z, z) = 0.$$

Show that d' is a metric on A, Lipschitz equivalent to d. Define $f: A \to A$ by $f(x) = y$, $f(y) = z$, $f(z) = z$. Show that f is a d'-contraction but not a d-contraction.

18*. Let $f: M \to M$ be a map of a compact metric space such that for any x, y in M with $x \neq y$, $d(f(x), f(y)) < d(x, y)$. Prove that f has a unique fixed point. (Hint: show that $\inf\{d(x, f(x)) : x \in M\}$ is attained and get a contradiction unless this inf is zero.)

19. Give examples of
    (a) a nested sequence of non-empty closed sets $V_n \subset (0, 1)$ with diam $V_n \to 0$ as $n \to \infty$ but $\bigcap_{n=1}^{\infty} V_n = \varnothing$,

(b) a nested sequence of non-empty sets U_n in $\mathbf{R}$ with diam $U_n \to 0$ as $n \to \infty$, but $\bigcap_{n=1}^{\infty} U_n = \varnothing$,

(c) a nested sequence of non-empty closed $V_n \subset \mathbf{R}$ with $\bigcap_{n=1}^{\infty} V_n = \varnothing$.

20. Give an example of two everywhere dense sets in a complete metric space whose intersection is empty.

21. Show that there is no metric on $\mathbf{Q}$ which is topologically equivalent to the usual metric and which makes $\mathbf{Q}$ complete. (Hint: use Baire's theorem.)

22*. Show that $\mathbf{Q}$ is not the intersection of a sequence of sets (U_n) where each U_n is open in $\mathbf{R}$. (Hint: consider $(\mathbf{R} - U_n) \cup \{q_n\}$ where (q_n) is an enumeration of $\mathbf{Q}$.)

10

Criteria for compactness in metric spaces

10.1 A general criterion

THE aim of this chapter is to exhibit interesting examples of compact metric spaces outside of Euclidean spaces. This section gives a general criterion for compactness of a metric space, and § 10.2 applies it to prove the Arzelà–Ascoli theorem on compactness for subspaces of function spaces with the sup metric.

We have already seen that a compact metric space is complete (Proposition 9.2.4) and totally bounded (Proposition 7.2.9). Completeness and total boundedness together imply compactness for a metric space. Before proving this, we discuss total boundedness in more detail. Recall that a metric space M is totally bounded if for any $\varepsilon>0$ there exists a finite ε-net for M.

LEMMA 10.1.1. *A subspace M' of a metric space M is totally bounded iff for any $\varepsilon>0$ there exists a finite subset $S\subset M$ such that $M'\subset \bigcup_{x\in S} B_\varepsilon(x)$.*

(The force of the lemma is that S need not be a subset of M'.)

Proof. Suppose that the stated condition holds and let $\varepsilon>0$. Applying the condition for $\frac{1}{2}\varepsilon$, we get a finite subset $S=\{x_1, x_2, \ldots, x_n\}\subset M$ such that $M'\subset \bigcup_{i=1}^{n} B_{\frac{1}{2}\varepsilon}(x_i)$. We may assume that for each $i=1, 2, \ldots, n$, $B_{\frac{1}{2}\varepsilon}(x_i)\cap M'\neq\varnothing$, otherwise x_i may be thrown out of S and we should still have $M'\subset \bigcup_{x\in S-\{x_i\}} B_{\frac{1}{2}\varepsilon}(x)$. For each $i=1, 2, \ldots, n$, choose y_i in $B_{\frac{1}{2}\varepsilon}(x_i)\cap M'$. Then $\{y_1, y_2, \ldots, y_n\}$ is an ε-net for M', since given y in M', $d(y, x_i)<\frac{1}{2}\varepsilon$ for some i in $\{1, 2, \ldots, n\}$, and

$$d(y, y_i)\leq d(y, x_i)+d(x_i, y_i)<\varepsilon.$$

COROLLARY 10.1.2. *A subspace of a totally bounded metric space is totally bounded.*

PROPOSITION 10.1.3. *If a subspace M' of a metric space M is totally bounded then so is $Cl(M')$.*

Proof. Let $\varepsilon > 0$, and let $S = \{x_1, x_2, \ldots, x_n\}$ be a $\frac{1}{2}\varepsilon$-net for M'. Then S is an ε-net for $Cl(M')$, since given y in $Cl(M')$, there exist x in M' with $d(x, y) < \frac{1}{2}\varepsilon$, and i in $\{1, 2, \ldots, n\}$ such that $d(x, x_i) < \frac{1}{2}\varepsilon$, so $d(y, x_i) < \varepsilon$.

PROPOSITION 10.1.4. *Any totally bounded metric space M is bounded.*

Proof. Suppose that $\{x_1, x_2, \ldots, x_n\}$ is, say, a 1-net for M. Then $M = \bigcup_{i=1}^{n} B_1(x_i)$. Since each $B_1(x_i)$ is bounded and a finite union of bounded sets is bounded (Proposition 2.2.15), M is bounded.

PROPOSITION 10.1.5. *Any bounded subset of $\mathbf{R}^n$ is totally bounded.*

Proof. This may be proved directly (using Exercise 10.3.1) or as follows. If $A \subset \mathbf{R}^n$ is bounded, then $Cl(A)$ is closed and bounded and hence compact. Then by Proposition 7.2.9, $Cl(A)$ is totally bounded and by Corollary 10.1.2, A is totally bounded.

As one might guess from this proof, the result can break down outside of Euclidean spaces.

EXAMPLE 10.1.6. Let $\mathscr{C}[0, 1]$ be the space of continuous real-valued functions on $[0, 1]$ with the sup metric d_∞. Let (f_n) be the sequence in Example 8.2.8 (the moving bump). For each n, $d_\infty(f_n, 0) = 1$, where 0 denotes the zero function. Thus $M = \{f_n : n \in \mathbf{N}\}$ is bounded. However, $d_\infty(f_n, f_m) = 1$ for any m, n with $m \neq n$, and there is therefore no finite 1-net in M. Thus M is not totally bounded.

LEMMA 10.1.7. *Let M be a totally bounded metric space. Then any sequence (x_n) in M has a Cauchy subsequence.*

Proof. This is a generalized bisection argument. Since M is totally bounded, it is covered by a finite number of open $\frac{1}{2}$-balls. At least one of these balls must contain x_n for infinitely many values of n. Hence there is a subsequence, say $(x_{n,1})$, of (x_n) such that all the

$x_{n,1}$ belong to a single open $\frac{1}{2}$-ball, so $d(x_{n,1}, x_{m,1}) < 1$ for all m, n. Suppose inductively that for $i = 1, 2, \ldots, k-1$, there exists a subsequence $(x_{n,i})$ of (x_n) such that $d(x_{n,i}, x_{m,i}) < 1/i$ for all m, n, and $(x_{n,i})$ is a subsequence of $(x_{n,i-1})$ for $i = 2, 3, \ldots, k-1$. Since M is covered by a finite number of open 1/2k-balls, at least one such ball contains $x_{n,k-1}$ for infinitely many values of n, so there is a subsequence $(x_{n,k})$ of $(x_{n,k-1})$ such that $d(x_{n,k}, x_{m,k}) < 1/k$ for all m, n. Now consider the 'diagonal' subsequence $(x_{n,n})$. For all $m \geqslant n$, $d(x_{n,n}, x_{m,m}) < 1/n$, since $x_{n,n}$ and $x_{m,m}$ both belong to the subsequence $x_{1,n}, x_{2,n}, x_{3,n}, \ldots$. Hence $(x_{n,n})$ is a Cauchy subsequence of (x_n) as required.

The converse of this lemma is also true (Exercise 10.3.4).

THEOREM 10.1.8. *A complete, totally bounded metric space M is compact.*

Proof. Let (x_n) be any sequence in M. By Lemma 10.1.7, (x_n) has a Cauchy subsequence, and by completeness of M, this subsequence converges. Hence M is compact.

This theorem explains the alternative name 'precompact' for totally bounded: in fact the completion (in the sense of the appendix) of a precompact metric space is compact.

COROLLARY 10.1.9. *The closure of a totally bounded subspace of a complete metric space is compact.*

This follows from Propositions 9.2.3(b), 10.1.3 and Theorem 10.1.8.

*10.2 Arzelà–Ascoli theorem

Compact subspaces are harder to find in function spaces than in $\mathbf{R}^n$. For example in $\mathscr{C}[0, 1]$ the closed unit ball $Cl(B_1(0))$ is not compact. For, as we saw in § 10.1, it contains the moving bump sequence (f_n) which has no Cauchy (and hence no convergent) subsequence. The same argument shows that $Cl(B_\varepsilon(0))$ is not compact, no matter how small $\varepsilon > 0$ is (we have only to use a moving bump sequence in which each bump has height ε). We need some extra condition on a closed bounded subspace of $\mathscr{C}[0, 1]$ in order to get compactness. The Arzelà–Ascoli theorem tells us what the required extra condition is.

DEFINITION 10.2.1. A subset $\mathscr{F} \subset \mathscr{C}[0, 1]$ is *equicontinuous at a* if given $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x) - f(a)| < \varepsilon$ for all x in $B_\delta(a)$ and for all f in $\mathscr{F}$.

(For this section $B_\delta(a)$ means the open ball with respect to the Euclidean metric restricted to $[0, 1]$.)

The force of this definition is that given $\varepsilon > 0$, the same $\delta > 0$ has to do the job for all f in $\mathscr{F}$. Thus the condition is global in $\mathscr{F}$. A similar definition may be given for any set of functions from a topological space T to a metric space M.

EXAMPLE 10.2.2. Let $\mathscr{F}$ be the set of members of the moving bump sequence. Then for any a in $(0, 1]$, $\mathscr{F}$ is equicontinuous at a. For we may choose N so that $2^{-N} < \frac{1}{2}a$, and then $f_n(x) = 0$ for all $n > N$ and all x in $B_{\frac{1}{2}a}(a)$, since we have chosen N so that beyond N the bump has passed entirely to the left of $\frac{1}{2}a$. Clearly then $\mathscr{F}_N = \{f_n : n > N\}$ is equicontinuous at a: for given $\varepsilon > 0$, we may take $\delta = \frac{1}{2}a$ and equicontinuity is trivially fulfilled by $\mathscr{F}_N$. Now each f_i for $1 \leqslant i \leqslant N$ is continuous at a, and hence $\{f_i\}$ is an equicontinuous set. Since a finite union of sets equicontinuous at a is again equicontinuous at a (Exercise 10.3.6), it follows that $\mathscr{F}$ is equicontinuous at a.

However, $\mathscr{F}$ is not equicontinuous at 0. To see this, take $\varepsilon = 1$ and any $\delta > 0$. For each n set $x_n = 3.2^{-n}$. Then a look at (f_n) shows that $f_n(x_n) = 1$ for all n. Hence for n such that $x_n < \delta$, we have $|f_n(x_n) - f_n(0)| = 1$ and $|x_n - 0| < \delta$, violating the equicontinuity condition.

Just as for ordinary continuity, we have the following definition.

DEFINITION 10.2.3. A subset $\mathscr{F} \subset \mathscr{C}[0, 1]$ is *equicontinuous* if it is equicontinuous at every a in $[0, 1]$.

An analogous definition may be made for subsets of $\mathscr{C}(T, M)$ for any topological space T and metric space M.

Equicontinuity is sometimes taken to mean what we shall call uniform equicontinuity.

DEFINITION 10.2.4. A subset $\mathscr{F} \subset \mathscr{C}[0, 1]$ is *uniformly equicontinuous* if given $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for any f in $\mathscr{F}$ and any x, y in $[0, 1]$ satisfying $|x - y| < \delta$.

An analogous definition can be made for subsets of $\mathscr{C}(M, M')$ for any metric spaces M, M'. Any member of a uniformly equicontinuous $\mathscr{F} \subset \mathscr{C}(M, M')$ is automatically uniformly continuous on M.

Definition 10.2.4 is superficially stronger than Definition 10.2.3, since it is global in $[0, 1]$ as well as in $\mathscr{F}$. In the general context of $\mathscr{C}(M, M')$ it would be stronger. But by the compactness of $[0, 1]$, any equicontinuous $\mathscr{F} \subset \mathscr{C}[0, 1]$ is uniformly equicontinuous (Exercise 10.3.8).

EXAMPLE 10.2.5. Let $\mathscr{F} \subset \mathscr{C}[0, 1]$ consist of all those functions f which satisfy the Lipschitz condition $|f(x)-f(y)| \leq K\,|x-y|$ for all x, y in $[0, 1]$ and some fixed $K > 0$. Given $\varepsilon > 0$, take $\delta = \varepsilon/K$. Then for $|x-y| < \delta$ we have $|f(x)-f(y)| \leq K\,|x-y| < \varepsilon$ for all f in $\mathscr{F}$, and $\mathscr{F}$ is uniformly equicontinuous.

The connection between equicontinuity and compactness begins to emerge in the next proposition.

PROPOSITION 10.2.6. *If* $\mathscr{F} \subset \mathscr{C}[0, 1]$ *is totally bounded then it is equicontinuous.*

Proof. Let $a \in [0, 1]$. Given $\varepsilon > 0$, let $\{f_1, f_2, \ldots, f_n\} \subset \mathscr{F}$ be an $\varepsilon/3$-net for $\mathscr{F}$. Each f_i is continuous at a, so there exists $\delta_i > 0$ such that $|f_i(x)-f_i(a)| < \varepsilon/3$ for all x in $B_{\delta_i}(a)$. Let $\delta = \mathrm{Min}\{\delta_1, \delta_2, \ldots, \delta_n\}$. Let $x \in B_\delta(a)$ and $f \in \mathscr{F}$. Then there exists some i in $\{1, 2, \ldots, n\}$ such that $d_\infty(f, f_i) < \varepsilon/3$, so

$$|f(x)-f(a)| \leq |f(x)-f_i(x)| + |f_i(x)-f_i(a)| + |f_i(a)-f(a)| < \varepsilon,$$

as required.

The same proof works for a totally bounded $\mathscr{F} \subset \mathscr{C}(T, M)$-one has only to replace $B_{\delta_i}(a)$ by an open set U_i containing a and such that $f_i(U_i) \subset B_{\varepsilon/3}(f_i(a))$.

Since any compact metric space is totally bounded (Proposition 7.2.9) this shows that equicontinuity is a necessary condition for compactness of $\mathscr{F}$.

THEOREM 10.2.7 (Arzelà–Ascoli). *A closed bounded subset* $\mathscr{F} \subset \mathscr{C}[0, 1]$ *is compact iff it is equicontinuous.*

Proof. We have already proved this in one direction. Suppose that $\mathscr{F}$ is equicontinuous. Since $\mathscr{F}$ is a closed subspace of the complete space $\mathscr{C}[0, 1]$, $\mathscr{F}$ is complete. Compactness of $\mathscr{F}$ will therefore follow from Theorem 10.1.8 if we show that $\mathscr{F}$ is totally bounded.

Let $\varepsilon > 0$. We shall construct a finite ε-net for $\mathscr{F}$ in the space of all bounded real-valued functions on $[0, 1]$ (with the sup metric).

Total boundedness of $\mathscr{F}$ then follows from Lemma 10.1.1. Since $\mathscr{F}$ is equicontinuous and $[0, 1]$ is compact, $\mathscr{F}$ is uniformly equicontinuous (Exercise 10.3.8), so we may choose $\delta>0$ such that $|f(x)-f(y)|<\varepsilon/3$ for any f in $\mathscr{F}$ and any x, y in $[0, 1]$ with $|x-y|<\delta$. Let $\{x_1, x_2, \ldots, x_n\}$ be a δ-net for $[0, 1]$. For the purposes of a diagram, we may suppose that the x_i are in increasing order and equally spaced out in $[0, 1]$ (see Fig. 10.1). We may now regard $[0, 1]$ as a *disjoint* union of sets S_i, where $S_i \subset B_\delta(x_i)$, by letting $S_1 = B_\delta(x_1)$ and inductively putting $S_r = B_\delta(x_r) - \bigcup_{i=1}^{r-1} S_i$. We may suppose that each S_i is non-empty (otherwise just delete it). In the diagram, the S_i are the intervals $[0, x_1+\delta)$, $[x_1+\delta, x_2+\delta), \ldots, [x_{n-1}+\delta, 1]$. The point of considering this disjoint union is that we shall construct functions on $[0, 1]$ by specifying different constant values on the S_i. Since $\mathscr{F}$ is bounded, there exists Δ such that $d_\infty(f, 0) \leqslant \Delta$ for all f in $\mathscr{F}$, so $|f(x)| \leqslant \Delta$ for all f in $\mathscr{F}$ and all x in $[0, 1]$. Choose an $\varepsilon/3$-net $\{y_1, y_2, \ldots, y_m\}$ for $[-\Delta, \Delta]$. It is convenient in the diagram to make the y_i symmetric about 0 and equally spaced at $\varepsilon/3$ apart. Now consider the collection Ψ of all maps

$$\psi : \{1, 2, \ldots, n\} \rightarrow \{1, 2, \ldots, m\}.$$

(There are m^n such maps in all.) For each ψ in Ψ, we define $g_\psi : [0, 1] \rightarrow \mathbf{R}$ by letting g take the constant value $y_{\psi(i)}$ on S_i for each i. Thus each g_ψ is a 'step function' on the grid in Fig. 10.1. We aim to show that $\{g_\psi : \psi \in \Psi\}$ is an ε-net for $\mathscr{F}$.

Consider a fixed f in $\mathscr{F}$. For any $i = 1, 2, \ldots, n$, we have $f(x_i) \in B_{\varepsilon/3}(y_j)$ for at least one j. (In the diagram, there will be either one or two possible values of j for each pair (f, i).) Define ψ by putting, for each i, $\psi(i) = j$ where $f(x_i) \in B_{\varepsilon/3}(y_j)$. We shall prove that $d_\infty(f, g_\psi) \leqslant 2\varepsilon/3$. For given x in $[0, 1]$, $x \in S_i \subset B_\delta(x_i)$ for some unique i, and $g_\psi(x) = y_{\psi(i)}$ where $f(x_i) \in B_{\varepsilon/3}(y_{\psi(i)})$. Hence

$$|f(x) - g_\psi(x)| \leqslant |f(x) - f(x_i)| + |f(x_i) - g_\psi(x)| < 2\varepsilon/3,$$

so

$$d_\infty(f, g_\psi) = \sup_{x \in [0,1]} |f(x) - g_\psi(x)| \leqslant 2\varepsilon/3 < \varepsilon.$$

This proves that $\{g_\psi : \psi \in \Psi\}$ is an ε-net for $\mathscr{F}$ as required. In fact not all the g_ψ are needed. In the case of Fig. 10.1, the step functions g_ψ which the above procedure chooses cannot jump more than two vertical steps at a time. Figure 10.1 shows the graph

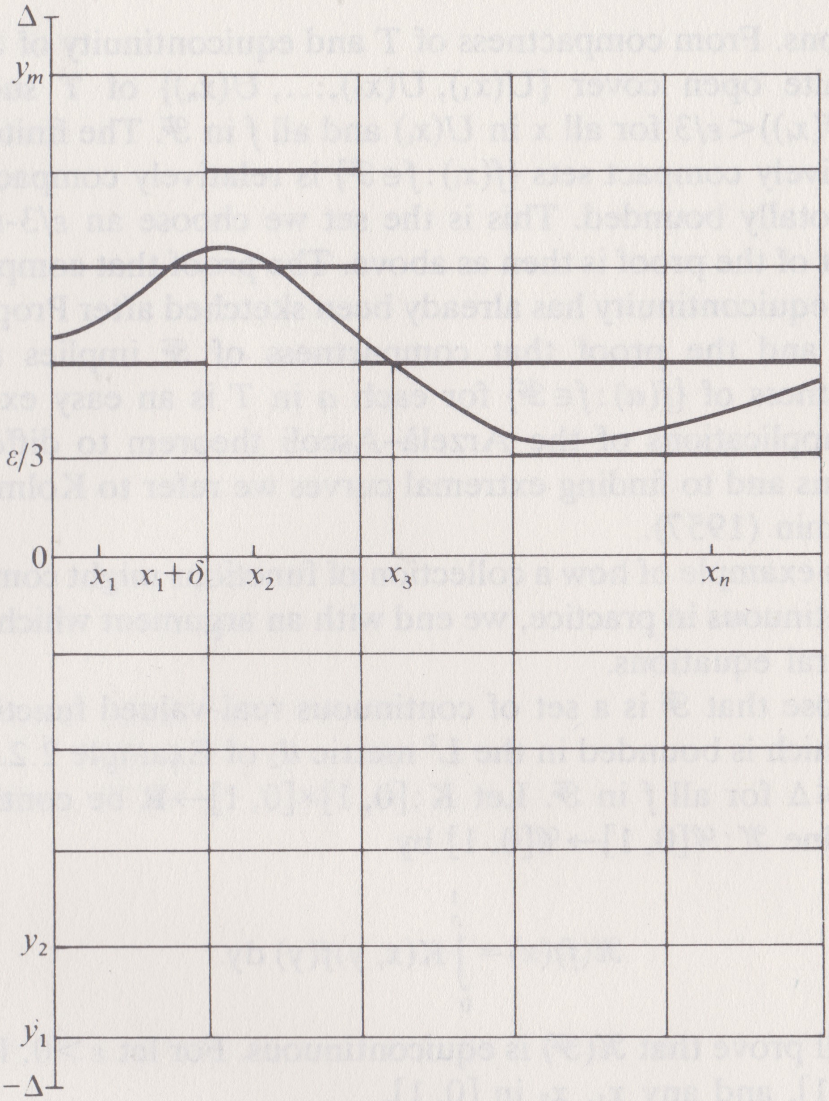


FIG. 10.1

of a function f in $\mathscr{F}$ together with all the possible steps which a corresponding g_ψ might take.

The above procedure for constructing the g_ψ is rather brutal, but it is fairly specific and also capable of generalization. For an alternative in which the ε-net lies in $\mathscr{C}[0, 1]$, see for example Dieudonné (1960). For a direct proof of sequential compactness of $\mathscr{F}$ see Copson (1968).

A more general (though not the most general) version of the theorem says that if T is a compact topological space and M is a complete metric space, then $\mathscr{F} \subset \mathscr{C}(T, M)$ is relatively compact (recall this means that $Cl(\mathscr{F})$ is compact) iff $\mathscr{F}$ is equicontinuous and for each a in T, $\{f(a) : f \in \mathscr{F}\}$ is relatively compact in M. The above proof needs only minor modifications to prove sufficiency of these conditions. From compactness of T and equicontinuity of $\mathscr{F}$ there is a finite open cover $\{U(x_1), U(x_2), \ldots, U(x_n)\}$ of T such that $d(f(x), f(x_i)) < \varepsilon/3$ for all x in $U(x_i)$ and all f in $\mathscr{F}$. The finite union of relatively compact sets $\{f(x_i) : f \in \mathscr{F}\}$ is relatively compact in M, hence totally bounded. This is the set we choose an $\varepsilon/3$-net for. The rest of the proof is then as above. The proof that compactness implies equicontinuity has already been sketched after Proposition 10.2.6, and the proof that compactness of $\mathscr{F}$ implies relative compactness of $\{f(a) : f \in \mathscr{F}\}$ for each a in T is an easy exercise.

For applications of the Arzelà–Ascoli theorem to differential equations and to finding extremal curves we refer to Kolmogorov and Fomin (1957).

As an example of how a collection of functions might come to be equicontinuous in practice, we end with an argument which occurs in integral equations.

Suppose that $\mathscr{F}$ is a set of continuous real-valued functions on $[0, 1]$ which is bounded in the L^2 metric d_2 of Example 2.2.11, say $d_2(f, 0) \leqslant \Delta$ for all f in $\mathscr{F}$. Let $K : [0, 1] \times [0, 1] \to \mathbf{R}$ be continuous, and define $\mathscr{K} : \mathscr{C}[0, 1] \to \mathscr{C}[0, 1]$ by

$$\mathscr{K}(f)(x) = \int_0^1 K(x, y) f(y)\, dy.$$

We shall prove that $\mathscr{K}(\mathscr{F})$ is equicontinuous. For let $\varepsilon > 0$. For any $f \in \mathscr{C}[0, 1]$, and any x_1, x_2 in $[0, 1]$,

$$\begin{aligned} |\mathscr{K}f(x_1) - \mathscr{K}f(x_2)| &= \left| \int_0^1 (K(x_1, y) - K(x_2, y)) f(y)\, dy \right| \\ &\leqslant \left\{ \int_0^1 (K(x_1, y) - K(x_2, y))^2\, dy \right\}^{\frac{1}{2}} \left\{ \int_0^1 (f(y))^2\, dy \right\}^{\frac{1}{2}}, \end{aligned}$$

by the C-S-B inequality for integrals. Now K is continuous on the compact space $[0, 1] \times [0, 1]$, so K is uniformly continuous, and in particular there exists $\delta > 0$ such that for $|x_1 - x_2| < \delta$, $|K(x_1, y) - K(x_2 - y)| < \varepsilon/2\Delta$ for all y in $[0, 1]$. Also, by hypothesis $\left(\int_0^1 f^2 \right)^{\frac{1}{2}} \leqslant \Delta$ for all f in $\mathscr{F}$. Thus for any f in $\mathscr{F}$ and for any x_1, x_2 *in* $[0, 1]$ satisfying $|x_1 - x_2| < \delta$, $|\mathscr{K}f(x_1) - \mathscr{K}f(x_2)| \leqslant \varepsilon/2 < \varepsilon$, so $\mathscr{K}(\mathscr{F})$ is uniformly equicontinuous.

Also, $\mathcal{K}(\mathcal{F})$ is bounded in the sup metric, since for all x in $[0, 1]$,

$$|\mathcal{K}f(x)| = \left| \int_0^1 K(x, y)f(y)\,dy \right|$$

$$\leqslant \left\{ \int_0^1 (K(x, y))^2\,dy \right\}^{\frac{1}{2}} \left\{ \int_0^1 (f(y))^2\,dy \right\}^{\frac{1}{2}} \leqslant \Delta'.\Delta,$$

where Δ' is an upper bound for the continuous function k on $[0, 1]$ given by

$$k(x) = \left\{ \int_0^1 (K(x, y))^2\,dy \right\}^{\frac{1}{2}}.$$

Thus $\mathcal{K}$ takes sets which are bounded in the L^2 metric into sets which are bounded and equicontinuous, and hence by the Arzelà–Ascoli theorem relatively compact in the sup metric.*

Exercises 10.3

1. Prove that a product metric space $M \times M'$ is totally bounded iff M and M' are totally bounded.
2. (a) Prove that if $f: M \to M'$ is a uniformly continuous function of metric spaces and if M is totally bounded then $f(M)$ is totally bounded. Deduce that total boundedness is an invariant under uniform equivalence of metrics.
   (b) Give an example to show that total boundedness is not a topological invariant.
3. Show that any totally bounded metric space is second countable (has a countable basis).
4. Show that if every sequence in a metric space M has a Cauchy subsequence then M is totally bounded. (Hint: this is really what the proof of Proposition 7.2.9 showed.)
5. Prove that ℓ^2 is not totally bounded.
6. Show that if for $i = 1, 2, \ldots, n$, $\mathcal{F}_i \subset \mathcal{C}[0, 1]$ is equicontinuous at a, then $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \ldots \cup \mathcal{F}_n$ is equicontinuous at a.
7. (i) Show that $\mathcal{F} = \{f_n : n \in \mathbf{N}\} \subset \mathcal{C}[0, 1]$ given by $f_n(x) = x^n$ for all x in $[0, 1]$, is equicontinuous at 0 but not at 1.
   (ii) Show that $\mathcal{F}$ in (i) is equicontinuous at every a in $(0, 1)$.

8. Prove that if $\mathscr{F} \subset \mathscr{C}[0, 1]$ is equicontinuous then it is uniformly equicontinuous. (Hint: go through the proof of Proposition 5.8.2, inserting 'for all f in $\mathscr{F}$' occasionally.)

9. Is the $\mathscr{F}$ in Exercise 7 uniformly equicontinuous on $[0, 1)$?

10. Prove that if $\mathscr{F} \subset \mathscr{C}[0, 1]$ is equicontinuous and if for each a in $[0, 1]$, $\{f(a) : f \in \mathscr{F}\}$ is bounded, then $\mathscr{F}$ is bounded (in the sup metric).

11. Show that if $\mathscr{F} \subset \mathscr{C}[0, 1]$ is equicontinuous then the set of all a for which $V_a = \{f(a) : f \in \mathscr{F}\}$ is bounded, is open and closed in $[0, 1]$. Deduce that if $\mathscr{F}$ is equicontinuous and V_a is bounded for one point a in $[0, 1]$ then $\mathscr{F}$ is bounded.

12. Can you generalize Exercise 11 to equicontinuous subsets of $\mathscr{C}(T, M)$ where T is a topological space and M is a metric space?

13. Give an example of an equicontinuous $\mathscr{F} \subset \mathscr{C}[0, 1]$ which is not bounded.

14. Suppose that (f_n) is a sequence in $\mathscr{C}[0, 1]$ such that $\{f_n : n \in \mathbf{N}\}$ is equicontinuous. Prove that
$$A = \{a \in [0, 1] : (f_n(a)) \text{ is a Cauchy sequence}\}$$
is closed in $[0, 1]$.

15. Suppose that M_1, M_2, M_3 are metric spaces with M_1, M_2 compact, and that $F : M_1 \times M_2 \to M_3$ is continuous. For each x in M_1 let $F_x : M_2 \to M_3$ be given by $F_x(y) = F(x, y)$ for all y in M_2. Show that $\{F_x : x \in M_1\} \subset \mathscr{C}(M_2, M_3)$ is equicontinuous. (Hint: use uniform continuity of F.)

16*. Let M be a compact metric space and let $\mathscr{I} \subset \mathscr{C}(M, M)$ denote the set of all isometries of M (onto itself). Show that $\mathscr{I}$ is compact (in the sup metric). (Hint: use the general form of the Arzelà–Ascoli theorem.)

11

Appendix

11.1 Real numbers

In this section we sketch one way of constructing **R** from **Q**. Logically this has no advantages over the axiomatic approach of Chapter 1, because we still need axioms for **Q** (or for **N**). But psychologically it may help for two reasons. First, it is close to the common-sense approach to real numbers. Secondly, most of our examples of compact, connected or complete spaces rely on the theorems that $[a, b]$ is compact, connected and that **R** is complete. We now stand back a little from these theorems by avoiding taking as axiomatic an assertion (Axiom 1.1.4) from which they follow pretty immediately. Another motive for discussing this construction is its close connection with completion of metric spaces (see § 11.2).

We assume that **Q** is known. The reader interested in axioms for **N** and the subsequent construction of **Q** is referred to Landau (1951) and Mendelson (1964).

In practical calculations involving, say, $\sqrt{2}$ we use a rational number approximating it, probably a decimal expansion to n places. What is the number $\sqrt{2}$ itself? We might try defining it as the sequence (q_n), where q_n is the decimal expansion correct to n places: more precisely, the decimal $q_n = 1.a_1a_2\dots a_n$ making $|q_n^2-2|$ minimal among the numbers $|x^2-2|$ for all decimals x of the form $1.x_1x_2\dots x_n$. Then (q_n) is a Cauchy sequence since for $m>n$, $|q_n-q_m|<10^{-n}$. (We have to redefine Cauchy sequence: given any *rational* $\varepsilon>0$, there exists N such that $|q_n-q_m|<\varepsilon$ for $m, n\geqslant N$, and similarly for the definition of convergence.) However, there are disadvantages in defining a real number α to be a Cauchy sequence in **Q**. There are many different sequences in **Q** approximating α; even apart from ambiguities in decimal expansions, we could for example use a number other than 10 as the base. It is imprudent to tie ourselves down to one special approximating sequence as the *definition* of α. A related disadvantage appears when we try to define addition and multiplication in **R**. The natural way to add Cauchy sequences (q_n), (q'_n) gives $(q_n+q'_n)$ as their sum, but doing this with decimal expansions correct to n places may give the 'wrong' answer, although it gives another Cauchy sequence intuitively converging to the correct number. For example this way of calculating $\frac{2}{3}+\frac{2}{3}$ gives the sequence 1·4, 1·34, 1·334,... instead of the 'correct' 1·3, 1·33, 1·333,.... Similar remarks apply to multiplication.

It is therefore better to define a real number as an equivalence class of Cauchy sequences in **Q**, even although this is more sophisticated. The equivalence relation used is $(q_n)\sim(q'_n)$ iff $(q_n-q'_n)$ converges to zero, or in other words iff the sequences (q_n) and (q'_n) are intuitively approximating the same number. It turns out that the resulting equivalence classes, with fairly obvious definitions of addition and multiplication, satisfy the axioms alluded to in Chapter 1. Moreover, this construction is just a formulation of the practical viewpoint on real numbers, where we use different approximations to meet the needs of different situations. The reader who wants only to feel less mystery about **R** has now gone far enough.

†To get the algebraic and order properties of **R**, we assume the definitions of field, commutative ring, ring homomorphism, and the construction of the quotient ring of a commutative ring by an ideal to be known (see for example Herstein (1964)) but we define orderings since they are more central to what follows.

DEFINITION 11.1.1. An *order* for a field F is a subset P of F called the (strictly) positive elements, such that

(P1) $x, y\in P \Rightarrow x+y, xy\in P$.

(P2) For any x in F, exactly one of the following holds: $x\in P$, $x=0$, $-x\in P$.

Given an order P for a field F, we write $x>y$ (equivalently $y<x$) if $x-y\in P$, and $x\geqslant y$ (equivalently $y\leqslant x$) if $x>y$ or $x=y$.

We need some results about Cauchy sequences in **Q**. For convenience we call (q_n) a *null sequence* if it converges to zero, and a *non-null sequence* otherwise.

LEMMA 11.1.2. *Any Cauchy sequence in* **Q** *is bounded.*

This follows as in the proof of Theorem 1.2.9.

LEMMA 11.1.3. *If (q_n) is a non-null Cauchy sequence in* **Q** *then there exist (a rational number) $\delta>0$ and (an integer) N such that either*

(a) *$q_n>\delta$ for all $n\geqslant N$, or*

(b) *$q_n<-\delta$ for all $n\geqslant N$.*

Proof. The negation of '(q_n) converges to zero' says that for some (rational) $\varepsilon_0>0$ and any N there is some $M\geqslant N$ with $|q_M|\geqslant\varepsilon_0$. Since (q_n) is Cauchy, there exists N such that $|q_n-q_m|<\delta$ for $m, n\geqslant N$, where $\delta=\frac{1}{2}\varepsilon_0$. By choice of ε_0, there is some $M\geqslant N$ such that $|q_M|\geqslant 2\delta$. We now distinguish two cases.

(a) If $q_M\geqslant 2\delta$, then for any $n\geqslant N$, $q_M-q_n<\delta$ so $q_n>q_M-\delta\geqslant\delta$.

(b) If $q_M\leqslant -2\delta$, then for any $n\geqslant N$, $q_n-q_M<\delta$ so $q_n<q_M+\delta\leqslant-\delta$.

This completes the proof. In either case, $|q_n|>\delta$ for $n\geqslant N$.

Addition and multiplication of sequences in **Q** is defined by:

$$(q_n)+(q_n')=(q_n+q_n'), \qquad (q_n).(q_n')=(q_nq_n').$$

In particular if (q_n), (q_n') are Cauchy, then so are $(-q_n)$, (q_n+q_n'), (q_nq_n'). The proofs are easy. (Lemma 11.1.2 is needed for the product, and one should remember to talk about *rational* ε.) It also follows easily that the set $\mathscr{C}$ of all Cauchy sequences in **Q** forms a commutative ring, with multiplicative identity $e=(1, 1, 1,\ldots)$.

LEMMA 11.1.4. *The set $\mathscr{N}$ of all null sequences in* **Q** *forms an ideal in $\mathscr{C}$.*

One has to prove that if (q_n), $(q_n')\in\mathscr{N}$ and $(x_n)\in\mathscr{C}$ then (q_n-q_n'), $(q_nx_n)\in\mathscr{N}$. The proofs are easy, using Lemma 11.1.2 for the product.

Now **R** is defined as the quotient ring $\mathscr{C}/\mathscr{N}$. Thus a real number is an equivalence class of Cauchy sequences in **Q**, where $(q_n)\sim(q_n')$ iff (q_n-q_n') is null. It follows by algebra that **R** is a commutative ring, with multiplicative identity $\hat{1}$ the equivalence class of e. (For clarity we denote elements of **R** by $\hat{x}$, $\hat{y}$ etc. in this section.)

PROPOSITION 11.1.5. **R** *is a field.*

Proof. What remains to be proved is that if $\hat{x}\neq\hat{0}$ in **R**, then there exists $\hat{y}$ in **R** with $\hat{x}\hat{y}=\hat{1}$. Let (q_n) in $\mathscr{C}$ represents $\hat{x}$. By definition of **R**, we have to find (q_n') in $\mathscr{C}$ such that $(q_nq_n'-1)\in\mathscr{N}$. (For then $(q_nq_n')\sim e$, so $\hat{x}\hat{y}=1$, where $\hat{y}$ is the real number represented by (q_n').)

Since $\hat{x}\neq\hat{0}$, (q_n) is non-null and by Lemma 11.1.3 there exists $\delta>0$ and N such that $|q_n|>\delta$ for all $n\geqslant N$. Let $q'_n=1/q_n$ for $n\geqslant N$ and let $q'_n=0$ (say) for $n<N$. Since $q_nq'_n-1=0$ for $n\geqslant N$, $(q_nq'_n-1)$ is certainly a null sequence. It remains to show that (q'_n) is Cauchy. Given $\varepsilon>0$, let $N_1\geqslant N$ be such that $|q_n-q_m|<\varepsilon\delta^2$ for m, $n\geqslant N_1$. Then for m, $n\geqslant N_1$,

$$|q'_n-q'_m|=|q_m-q_n|/|q_nq_m|<|q_m-q_n|/\delta^2<\varepsilon,$$

as required.

Next we define $i:\mathbf{Q}\to\mathbf{R}$ by letting $i(q)$ be the equivalence class of the sequence $(q, q, q,\ldots)$. Formally, $i=\pi\circ j$, where $j:\mathbf{Q}\to\mathscr{C}$ is given by $j(q)=(q, q, q,\ldots)$ and $\pi:\mathscr{C}\to\mathbf{R}$ assigns to each element of $\mathscr{C}$ its equivalence class in $\mathbf{R}=\mathscr{C}/\mathscr{N}$. Since j and π are ring homomorphisms, so is i. The kernel of i is zero, since if $q\neq 0$, then $(q, q, q,\ldots)$ is non-null. Thus i determines a ring isomorphism of $\mathbf{Q}$ onto $i(\mathbf{Q})$. In practice $\mathbf{Q}$ is identified with $i(\mathbf{Q})$ via i, but for clarity in this section we preserve the distinction between them.

We now order $\mathbf{R}$ by defining $P\subset\mathbf{R}$ to be the set of those elements represented by Cauchy sequences satisfying condition (a) of Lemma 11.1.3. To show that P is well defined, we need to prove that if (q_n) in $\mathscr{C}$ satisfies 11.1.3(a) then so does $(q_n+q'_n)$ for any (q'_n) in $\mathscr{N}$. The proof is easy and we omit it. It is also easy to show that P satisfies the conditions for an order. ((P2) follows from Lemma 11.1.3.) Notice that for q in $\mathbf{Q}$, $i(q)\in P$ iff $q>0$, since $i(q)$ is represented by $(q, q, q,\ldots)$. Thus i is order-preserving. The definitions in § 1.2 of convergent and Cauchy sequences in $\mathbf{R}$ once more make sense, since we have defined an order for $\mathbf{R}$.

For convenience later, we discuss the order in $\mathbf{R}$ and its relation with the order in $\mathbf{Q}$ a little further.

LEMMA 11.1.6. *Suppose that* (q_n) *in* $\mathscr{C}$ *represents* $\hat{x}$ *in* $\mathbf{R}$. *If* $q_n\geqslant 0$ *for all sufficiently large* n, *then* $\hat{x}\geqslant\hat{0}$.

Proof. If $\hat{x}<\hat{0}$ then by definition of P, there would exist some $\delta>0$ such that $q_n<-\delta$ for all n sufficiently large, contradicting the hypothesis.

The converse is false; if $\hat{x}=\hat{0}$ then (q_n) is null, but this does not force $q_n\geqslant 0$ for all n sufficiently large: q_n might for example be $(-1)^n/n$.

COROLLARY 11.1.7. *Suppose that* (q_n) *represents* $\hat{x}$ *and that* $q_n\leqslant q$ $(q_n\geqslant q)$ *for all sufficiently large* n. *Then* $\hat{x}\leqslant i(q)$ $(\hat{x}\geqslant i(q))$.

This follows from the lemma since $(q-q_n)$ represents $i(q)-\hat{x}$ $(\pi((q-q_n))=\pi j(q)-\pi((q_n))=i(q)-\hat{x})$. The proof for $q_n\geqslant q$ is similar.

We also need to handle absolute values in $\mathbf{R}$.

LEMMA 11.1.8. *If (q_n) represents $\hat{x}$ then $(|q_n|)$ represents $|\hat{x}|$.*

Proof. If $\hat{x}=\hat{0}$ then (q_n) is a null sequence, hence so is $(|q_n|)$ and the latter represents $\hat{0}=|\hat{x}|$.

If $\hat{x}>\hat{0}$, so that $|\hat{x}|=\hat{x}$, then by definition of P, there exist $\delta>0$ and N such that $q_n>\delta$ for all $n\geqslant N$, so $|q_n|-q_n=0$ for $n\geqslant N$ and $(|q_n|-q_n)$ is null. Hence $(|q_n|)$ represents the same number as does (q_n), namely $\hat{x}=|\hat{x}|$.

If $\hat{x}<\hat{0}$, so that $|\hat{x}|=-\hat{x}$, then by definition of P, there exist $\delta>0$ and N such that $q_n<-\delta$ for all $n\geqslant N$. Thus for $n\geqslant N$, $|q_n|=-q_n$, so $(|q_n|+q_n)$ is null and $(|q_n|)$ represents the same real number as does $(-q_n)$, namely $-\hat{x}=|\hat{x}|$.

COROLLARY 11.1.9. *If (q_n) represents $\hat{x}$ and $|q_n|\leqslant q$ for all sufficiently large n then $|\hat{x}|\leqslant i(q)$.*

This follows from Lemma 11.1.8 and Corollary 11.1.7.

COROLLARY 11.1.10. *For any $\hat{x}$, $\hat{y}$ in $\mathbf{R}$, $|\hat{x}+\hat{y}|\leqslant|\hat{x}|+|\hat{y}|$.*

Proof. Let (q_n), (r_n) represent $\hat{x}$, $\hat{y}$. Then (q_n+r_n) represents $\hat{x}+\hat{y}$, so by Lemma 11.1.8, $(|q_n+r_n|)$ represents $|\hat{x}+\hat{y}|$. Similarly $(|q_n|)$, $(|r_n|)$ represent $|\hat{x}|$, $|\hat{y}|$. Thus $(|q_n|+|r_n|-|q_n+r_n|)$ represents $|\hat{x}|+|\hat{y}|-|\hat{x}+\hat{y}|$, and the result now follows from Lemma 11.1.6, and the triangle inequality for $\mathbf{Q}$, which we assume known.

Finally we need the following form of Archimedes' axiom.

PROPOSITION 11.1.11. *Given any $\hat{\varepsilon}>\hat{0}$ in $\mathbf{R}$, there exists δ in $\mathbf{Q}$ such that $\hat{0}<i(\delta)\leqslant\hat{\varepsilon}$. In particular there exists an integer n such that $i(1/n)\leqslant\hat{\varepsilon}$.*

Proof. Let (ε_n) represent $\hat{\varepsilon}$. Since $\hat{\varepsilon}>\hat{0}$, there exist $\delta>0$ and N such that $\varepsilon_n>\delta$ for $n\geqslant N$. Then $\hat{\varepsilon}\geqslant i(\delta)$ by Corollary 11.1.7, and $i(\delta)>\hat{0}$ since $\delta>0$. The last part of the statement follows since there is an integer $n>1/\delta$.

In particular it follows from this proposition that in defining convergent or Cauchy sequences in $\mathbf{R}$ it does not matter whether we use a 'rational or real ε' (precisely, whether $\hat{\varepsilon}$ is in $i(\mathbf{Q})$ or not).

The cheap trick by which $\mathbf{R}$ is forced by definition to be complete is more or less exposed in the next proposition.

PROPOSITION 11.1.12. *If (q_m) represents $\hat{x}$, then $(i(q_m))$ converges to $\hat{x}$ in* **R**.

Proof. Given any $\hat{\varepsilon} > \hat{0}$, by Lemma 11.1.11 there exists a rational $\delta > 0$ *with* $\hat{0} < i(\delta) \leqslant \hat{\varepsilon}$. Since $(q_n) \in \mathscr{C}$, there exists N such that $|q_n - q_m| < \delta$ for all $m, n \geqslant N$. Now think of a fixed $m \geqslant N$, and let n vary: the resulting sequence $(|q_n - q_m|)$ represents $|\hat{x} - i(q_m)|$ by Lemma 11.1.8. Hence by Corollary 11.1.9, $|\hat{x} - i(q_m)| \leqslant i(\delta) \leqslant \hat{\varepsilon}$. Since this is true for all $m \geqslant N$, $(i(q_m))$ converges to $\hat{x}$.

COROLLARY 11.1.13. *$i(\mathbf{Q})$ is dense in* **R**.

This follows from the proposition and Lemma 7.2.2.

THEOREM 11.1.14. *Any Cauchy sequence $(\hat{x}_n)$ in **R** converges to a point $\hat{x}$ in* **R**.

Proof. Since $i(\mathbf{Q})$ is dense in **R**, for each n there is some q_n in **Q** such that $|\hat{x}_n - i(q_n)| < i(1/n)$. We prove that $(q_n) \in \mathscr{C}$. For given (rational) $\delta > 0$, choose an integer N such that $|\hat{x}_n - \hat{x}_m| < i(\delta/3)$ for $m, n \geqslant N$, and also $N > 3/\delta$. Then for $m, n \geqslant N$, using Corollary 11.1.10,

$$|i(q_m) - i(q_n)| \leqslant |i(q_m) - \hat{x}_m| + |\hat{x}_m - \hat{x}_n| + |\hat{x}_n - i(q_n)| < i(\delta).$$

Hence $|q_m - q_n| < \delta$ for all $m, n \geqslant N$, and $(q_n) \in \mathscr{C}$. Thus (q_n) represents some $\hat{x}$ in **R**. We prove that $(\hat{x}_n)$ converges to $\hat{x}$. Let $\hat{\varepsilon} > \hat{0}$. Since, by Proposition 11.1.12, $(i(q_n))$ converges to $\hat{x}$, we may choose N so that for $n \geqslant N$, $|i(q_n) - \hat{x}| < i(\tfrac{1}{2})\hat{\varepsilon}$, and also (using Proposition 11.1.11) so that $i(1/N) < i(\tfrac{1}{2})\hat{\varepsilon}$. Then for $n \geqslant N$,

$$|\hat{x}_n - \hat{x}| \leqslant |\hat{x}_n - i(q_n)| + |i(q_n) - \hat{x}| < i(1)\hat{\varepsilon} = \hat{1} \,.\, \hat{\varepsilon} = \hat{\varepsilon},$$

as required.

Unfortunately, Theorem 11.1.14 does not give completeness in the form of Axiom 1.1.4. We therefore sketch the proof of Theorem 1.2.10 for **R** constructed as above.

From now on real numbers drop their hats and q in **Q** is identified with $i(q)$ in **R**.

Let S be a non-empty subset of **R** which is bounded above. Let $a \in S$ and let u be an upper bound for S. Then we take a look at $\tfrac{1}{2}(a+u)$, If this is an upper bound for S we set $a_2 = a$ and $u_2 = \tfrac{1}{2}(a+u)$. If not, we let $a_2 \in S \cap (\tfrac{1}{2}(a+u), \infty)$ and let $u_2 = u$. Carrying on inductively, we produce sequences (a_n), (u_n) such that $a_{n-1} \leqslant a_n \leqslant u_n \leqslant u_{n-1}$ and $|u_n - a_n| \leqslant (u-a)/2^n$. By Archimedes' axiom,

$(u-a)/2^n$ can be made smaller than any prescribed positive number by choosing n sufficiently large. It is easy to check that (a_n) and (u_n) are Cauchy sequences, hence convergent by Theorem 11.1.14, and that they converge to the same real number U, say. By construction, each u_n is an upper bound for S. If $x > U$ for some x in S, we should get the contradiction that $x > u_n$ for all n sufficiently large. Hence U is an upper bound for S. Similarly, no $x < U$ is an upper bound for S since $a_n > x$ for all n sufficiently large. Hence U is a least upper bound for S.

There are other ways of constructing **R** from **Q** (notably Dedekind cuts), but it can be shown that **R** is essentially unique, in that any two systems satisfying the axioms for **R** alluded to in Chapter 1 are isomorphic, by an isomorphism preserving order.†

11.2 Completion of metric spaces

The process of completing an arbitrary metric space is similar to the construction of **R** from **Q** in § 11.1, but it needs separate treatment, since the very definition of a metric assumes that **R** is already there, and completeness of **R** is used in completing a metric space.

DEFINITION 11.2.1. Given a metric space M with metric d, a *completion* of M consists of a metric space $\hat{M}$ with metric $\hat{d}$ together with a map $i: M \to \hat{M}$ such that

(C1) $\hat{M}$ is complete,

(C2) i is an isometry into, i.e. $\hat{d}(i(x), i(y)) = d(x, y)$ for all x, y in M,

(C3) $i(M)$ is dense in $\hat{M}$.

(C3) is there to make $\hat{M}$ minimal with respect to (C1) and (C2). For example if $M = \mathbf{Q}$ then $\mathbf{R}^2$ would satisfy (C1) and (C2), with $i: \mathbf{Q} \to \mathbf{R}^2$ defined by $i(q) = (q, 0)$, but **R** is a more economical completion of **Q**. Thus (C3) is needed if there is to be any hope of getting uniqueness of completions.

THEOREM 11.2.2. *Any metric space M with metric d has a completion.*

Proof. Let $\hat{M}$ be the set of equivalence classes of Cauchy sequences in M, where $(x_n) \sim (x'_n)$ iff $d(x_n, x'_n) \to 0$ as $n \to \infty$. It is clear that this is an equivalence relation. If $\hat{x}$, $\hat{y}$ in $\hat{M}$ are represented by the Cauchy sequences (x_n), (y_n), let $\hat{d}(\hat{x}, \hat{y}) = \lim_{n \to \infty} d(x_n, y_n)$. We first check that this limit exists by showing that $(d(x_n, y_n))$ is a Cauchy sequence in $\mathbf{R}$. For

$$|d(x_n, y_n)-d(x_m, y_m)|\leqslant d(x_m, x_n)+d(y_m, y_n)$$

by Exercise 2.6.3, so $(d(x_n, y_n))$ is a Cauchy sequence in $\mathbf{R}$ since (x_n) and (y_n) are Cauchy sequences in M. Hence, by completeness of $\mathbf{R}$, $\lim\limits_{n\to\infty} d(x_n, y_n)$ exists. If (x'_n) is another Cauchy sequence in M representing $\hat{x}$, then $d(x_n, x'_n)\to 0$ as $n\to\infty$, and

$$0\leqslant|d(x'_n, y_n)-d(x_n, y_n)|\leqslant d(x'_n, x_n)$$

by Exercise 2.6.2, so $\lim\limits_{n\to\infty} d(x'_n, y_n)=\lim\limits_{n\to\infty} d(x_n, y_n)$. It follows that if we change the representative Cauchy sequence for $\hat{y}$ also the limit still remains unchanged. Thus $\hat{d}$ is well defined.

Next we check that $\hat{d}$ is a metric for $\hat{M}$. Let $\hat{x}$, $\hat{y}$ in $\hat{M}$ be represented by (x_n), (y_n). We have

$$\hat{d}(\hat{x}, \hat{y})=0 \Leftrightarrow \lim_{n\to\infty} d(x_n, y_n)=0 \Leftrightarrow (x_n)\sim(y_n) \Leftrightarrow \hat{x}=\hat{y},$$

and clearly $\hat{d}(\hat{x}, \hat{y})\geqslant 0$, so (M1) holds. (M2) and (M3) follow (by taking limits) from (M2) and (M3) for d.

Let $i: M\to\hat{M}$ be defined by $i(x)=\hat{x}$ where $\hat{x}$ is represented by the Cauchy sequence $(x, x, x, \ldots)$ in M. Then

$$\hat{d}(i(x), i(y))=\lim_{n\to\infty} d(x, y)=d(x, y).$$

Thus i defines an isometry of M onto $i(M)$. As in § 11.1 we preserve the distinction between them.

The proofs that $i(M)$ is dense in $\hat{M}$ and that $\hat{M}$ is complete are analogous to (but easier than) the corresponding parts of § 11.1, and we do not repeat them.

In order to prove uniqueness we need a lemma.

LEMMA 11.2.3. *Given any dense subspace X of a metric space M_1 and an isometry $f: X\to M_2$ of X into a complete metric space M_2, f extends uniquely to an isometry of M_1 into M_2.*

Proof. Let d_1, d_2 denote the metrics in M_1, M_2. Given any x in M_1, there exists some sequence (x_n) in X converging to x, by denseness of X in M_1 (Lemma 7.2.2). Since f is an isometry into, $(f(x_n))$ is a Cauchy sequence in M_2 since (x_n) is a Cauchy (even convergent) sequence in M_1. Hence by completeness of M_2, $(f(x_n))$ converges to a point, say $g(x)$, in M_2. If (y_n) is another sequence in

X converging to x in M_1, then $d_1(x_n, y_n) \to 0$ as $n \to \infty$, and since f is an isometry, $d_2(f(x_n), f(y_n)) \to 0$ as $n \to \infty$, so $\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} f(y_n)$. Thus $g : M_1 \to M_2$ is well defined. In particular if $x \in X$ then (x_n) may be taken to be $(x, x, x, \ldots)$, so $g(x) = \lim_{n\to\infty} f(x) = f(x)$. Thus g is an extension of f. Moreover, if $x, y \in M_1$ and (x_n), (y_n) are sequences in X converging to x, y then

$$\begin{aligned} d_2(g(x), g(y)) &= d_2\left(\lim_{n\to\infty} f(x_n), \lim_{n\to\infty} f(y_n)\right) \\ &= \lim_{n\to\infty} d_2(f(x_n), f(y_n)) \\ &= \lim_{n\to\infty} d_1(x_n, y_n) \\ &= d_1\left(\lim_{n\to\infty} x_n, \lim_{n\to\infty} y_n\right) \\ &= d_1(x, y). \end{aligned}$$

Thus g is an isometry into. To prove uniqueness, note that if g is an isometry into, extending f, then g is continuous, so if (x_n) is a sequence in X converging to x in M_1, then $g(x) = \lim_{n\to\infty} g(x_n) = \lim_{n\to\infty} f(x_n)$.

From Lemma 11.2.3 we get the following 'universal property' of completions.

PROPOSITION 11.2.4. *Let $(\hat{M}, i)$ be a completion of the metric space M, and let i' be an isometry of M into a complete metric space M'. Then there is a unique isometry j of $\hat{M}$ into M' such that $j \circ i = i'$.*

Proof. Apply Lemma 11.2.3 to the dense subspace $i(M)$ of $\hat{M}$ and the isometry $i' \circ i^{-1}$ of $i(M)$ into M'.

The conclusion of Proposition 11.1.4 may be summed up in the following commutative diagram of isometries into:

$$\begin{tikzcd} M \arrow[r, "i"] \arrow[dr, "i'"'] & \hat{M} \arrow[d, "j"] \\ & M' \end{tikzcd}$$

Essential uniqueness of completions may now be proved.

THEOREM 11.2.5. *Suppose that $(\hat{M}, i)$ and (M', i') are completions of a metric space M. Then there is a unique isometry j of $\hat{M}$ onto M' such that $j \circ i = i'$.*

Proof. This may be followed using the diagram:

$$\begin{tikzcd}
M \arrow[r, "i"] \arrow[dr, "i'"'] & \hat{M} \arrow[d, shift right, "j"'] \\
 & M' \arrow[u, shift right, "j'"']
\end{tikzcd}$$

By Proposition 11.2.4, there is a unique isometry j of $\hat{M}$ into M' such that $j \circ i = i'$. Interchanging the roles of $\hat{M}$ and M', there is a unique isometry j' of M' into $\hat{M}$ such that $j' \circ i' = i$. Finally we apply Proposition 11.2.4 to the apparently trivial case where $M' = \hat{M}$, $i' = i$, to deduce that there is a unique isometry j'' of M into itself such that $j'' \circ i = i$. Since the identity map $1(\hat{M})$ of $\hat{M}$ obviously has this property, by uniqueness, $j'' = 1(\hat{M})$. But also $j' \circ j \circ i = j' \circ i' = i$, so by uniqueness of j'', $j' \circ j = 1(\hat{M})$. Similarly $j \circ j' = 1(M')$. Thus j and j' are mutually inverse isometries onto. Since also j carries the copy $i(M)$ of M in $\hat{M}$ isometrically onto the copy $i'(M)$ in M', completions are as unique as one could hope for.

There is another universal property possessed by completions, which we state in a sequence of results similar to the above, but without proofs. (The proofs are not hard.)

LEMMA 11.2.6. *Given a dense subspace X of a metric space M_1 and a uniformly continuous map $f : X \to M_2$ into a complete metric space M_2, f extends uniquely to a uniformly continuous map $g : M_1 \to M_2$.*

PROPOSITION 11.2.7. *Given any uniformly continuous map i' of M into a complete metric space M', there is a unique uniformly continuous map $j : \hat{M} \to M'$ such that $j \circ i = i'$ (where $(\hat{M}, i)$ is as before a completion of M).*

THEOREM 11.2.8. *Given two completions $(\hat{M}, i)$, (M', i') of M, there is a unique uniform equivalence $j : \hat{M} \to M'$ such that $j \circ i = i'$.*

By Theorem 11.2.5, j is in fact an isometry. The remaining force of Theorem 11.2.8 is that j is unique even within the wider class of uniform equivalences.

Lest these two different universal properties of completions (Propositions 11.2.4 and 11.2.7) should seem mysterious, it should be mentioned that we have suppressed in this book a whole class of spaces, called uniform spaces, which are more general than metric spaces and less general than topological spaces. There is a completion process for uniform spaces too, and the universal property in

Proposition 11.2.7 is really that possessed by the completion of the uniform space 'underlying' the given metric space.

As a very particular application of uniqueness, consider the process of completing the open interval $(0, 1)$. The general machinery builds up a sophisticated space in which the points are equivalence classes of Cauchy sequences in $(0, 1)$. But we can also see immediately that $[0, 1]$ and the inclusion map $i:(0, 1)\rightarrow[0, 1]$ satisfy the definition of completion for $(0, 1)$. Uniqueness tells us that this is indeed the completion given by the general machinery (up to isometry).

12

Guide to exercises

THIS contains hints, remarks, answers, sketch solutions—in telegraphese. Sketch solutions are not claimed to be unique or best possible. It is hoped that no peeking at this section will occur until serious efforts have been made on the exercises.

Exercises 1.5

5. Suppose $m/n=(p/q)^2$, $p, q\in\mathbf{Z}$ and h.c.f. $(p, q)=1$. Then $mq^2=np^2$, so $p^2 \mid mq^2$ (p^2 divides mq^2). Hence $p^2 \mid m$ since $(p, q)=1$. Thus $m=p^2k$ for $k\in\mathbf{Z}$. Then $kq^2p^2=np^2$, so $kq^2=n$. Hence $k \mid n$, contradicting $(m, n)=1$ unless $k=1$. So $n=q^2$ and $m=p^2$.

7. Depends on existence of *some* irrational, say $\sqrt{2}$. Given $x<y$ in $\mathbf{R}$, by 1.1.7 there exists q in $\mathbf{Q}$ with $x<q<y$ and then q' in $\mathbf{Q}$ with $q<q'<y$. Since $q'-q>0$, $\sqrt{2}/(q'-q)\in\mathbf{R}$ and there exists an integer $n>\sqrt{2}/(q'-q)$. Thus $x<q<q+(\sqrt{2}/n)<q'<y$, and $q+(\sqrt{2}/n)\notin\mathbf{Q}$.

11. Use $r^{1/n}\to 1$ as $n\to\infty$ (follows e.g. from $0<r^{1/n}<n^{1/n}$ for $n>r$, and Exercise 10).

13. First, $x_2=2/(x_1+1)>1$. Also,

$$x_3=2/(x_2+1)=2(x_1+1)/(3+x_1)>(2x_1+2)/4>x_1.$$

Similarly $x_4<x_2$. Suppose inductively that for some $n\geqslant 1$, $x_{2n}>1$, $x_{2n-1}<1$, $x_{2n+2}<x_{2n}$, $x_{2n+1}>x_{2n-1}$. Then $x_{2n+1}<1$ since $x_{2n}>1$, and similarly $x_{2n+2}>1$. Also, $x_{2n+3}=2/(x_{2n+2}+1)>2/(x_{2n}+1)=x_{2n+1}$. Similarly $x_{2n+4}<x_{2n+2}$. So (x_{2n}) monotonic decreasing, bounded below by 1, so converges to $a\geqslant 1$. Similarly (x_{2n+1}) converges to a' with $0<a'\leqslant 1$. From $x_{2n+1}=2/(x_{2n}+1)$ get $a'=2/(a+1)$. Similarly $a=2/(a'+1)$ so $a=2(a+1)/(3+a)$, whence $a=1$ or -2. But $a\geqslant 1$, so $a=1$. Then $a'=1$, and result follows from Exercise 12.

14. Since $x_1=\sqrt{2}$, $x_2>\sqrt{2}=x_1$. Suppose inductively that $x_n>x_{n-1}$

for some $n \geqslant 2$. Then

$$x_{n+1} = \{2+(x_n)^{\frac{1}{2}}\}^{\frac{1}{2}} > \{2+(x_{n-1})^{\frac{1}{2}}\}^{\frac{1}{2}} = x_n.$$

Also, $\sqrt{3} < x_n < 2$ for all $n \geqslant 2$ (by induction). So (x_n) converges to k, with $\sqrt{3} < k \leqslant 2$. From iteration formula get $k^4 - 4k^2 - k + 4 = 0$. Graph-sketching shows only one root $k > 1$.

21. Given rational $p/q \neq 0$, let $\varepsilon = 1/q$. For any $\delta > 0$, there is an irrational x with $|p/q - x| < \delta$, and then $|f(x) - f(p/q)| = 1/q$. Hence f not continuous at p/q.

    Let $a = 0$ or $a \notin \mathbf{Q}$, and let $\varepsilon > 0$. Choose $N > 1/\varepsilon$. There are only finitely many rationals $\{q_1, q_2, \ldots, q_n\}$, say in $[a-1, a+1]$, with denominators (when in lowest terms) $\leqslant N$. Let $\delta = \operatorname{Min}\{|a - q_i|\}$. For x with $|x - a| < \delta$, either $x \in \mathbf{Q}$ with denominator $> N$, and $|f(x) - f(a)| < 1/N < \varepsilon$, or $x = 0$ or $x \notin \mathbf{Q}$, and $|f(x) - f(a)| = 0 < \varepsilon$.

22. There are naive and sophisticated solutions. Here is a naive, geometric argument, using freely that convexity says graph f convex. Let $a < c < b$, and prove continuity at c: let L be straight line through points $(c, f(c))$, $(b, f(b))$ and L' the one through $(a, f(a))$, $(c, f(c))$. Then graph f is trapped in the cone formed by L, L' (argue this by convexity). See Fig. 12.1. Given $\varepsilon > 0$, choose $\delta < \varepsilon/\tau$ where $\tau = \operatorname{Max}\{|\tan\theta_1|, |\tan\theta_2|\}$. Then $|f(x) - f(c)| \leqslant \tau\,|x - c| < \varepsilon$ for $|x - c| < \delta$.

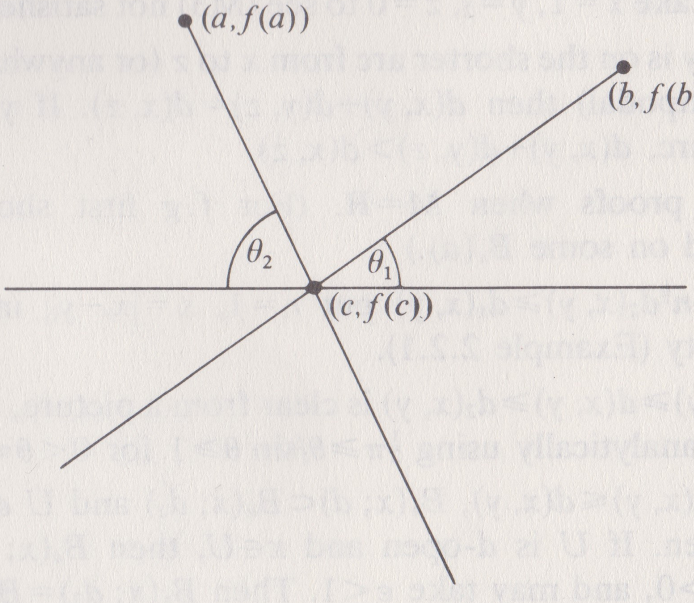


FIG. 12.1

Exercises 2.6

3.
$$|d(x, y)-d(z, t)| \leqslant |d(x, y)-d(y, z)|+|d(y, z)-d(z, t)|$$
$$\leqslant d(x, z)+d(y, t) \quad \text{by Exercise 2.}$$

In $\mathbf{R}^2$ this says various things about quadrilaterals, e.g. the difference in lengths of the diagonals is not greater than the sum of lengths of two opposite sides.

6. (M3) for d_2: if $d(x, y) \leqslant 1$, $d(y, z) \leqslant 1$ then
$$d_2(x, z) \leqslant d(x, z) \leqslant d(x, y)+d(y, z) = d_2(x, y)+d_2(y, z).$$
If $d(x, y) > 1$, then
$$d_2(x, z) \leqslant 1 = d_2(x, y) \leqslant d_2(x, y)+d_2(y, z),$$
and similarly if $d(y, z) > 1$. (M3) for d_3: first note that $f(a) = a/(1+a)$ is an increasing function of a for $a \geqslant 0$ (since $f'(a) > 0$). Next, if $a, b, c \geqslant 0$ and $a+b \geqslant c$, then
$$a/(1+a)+b/(1+b) = (a+b+2ab)/(1+a+b+ab)$$
$$\geqslant d/(1+d)$$
where $d = a+b+2ab \geqslant c$, so
$$a/(1+a)+b/(1+b) \geqslant c/(1+c),$$
and (M3) follows.

If M is discrete, $d_4 = d$ is a metric. If $M = \mathbf{R}$ with usual metric, take $x=1$, $y=\frac{1}{2}$, $z=0$ to see (M3) not satisfied by d_4.

8. (M3) If y is on the shorter arc from x to z (or anywhere on A if x, z antipodal) then $d(x, y)+d(y, z) = d(x, z)$. If y is on the longer arc, $d(x, y)+d(y, z) > d(x, z)$.

14. Imitate proofs when $M = \mathbf{R}$. (For $f.g$ first show f (say) bounded on some $B_\varepsilon(a)$.)

18. To get $n^{\frac{1}{2}}d_2(x, y) \geqslant d_1(x, y)$ put $r_i = 1$, $s_i = |x_i - y_i|$ in Cauchy's inequality (Example 2.2.1).

19. $\frac{1}{2}\pi d_2(x, y) \geqslant d(x, y) \geqslant d_2(x, y)$ is clear from a picture, and can be proved analytically using $\frac{1}{2}\pi \geqslant \theta/\sin\theta \geqslant 1$ for $0 < \theta \leqslant \frac{1}{2}\pi$.

20. Since $d_2(x, y) \leqslant d(x, y)$, $B_\varepsilon(x; d) \subset B_\varepsilon(x; d_2)$ and U d_2-open $\Rightarrow$ U d-open. If U is d-open and $x \in U$, then $B_\varepsilon(x; d) \subset U$ for some $\varepsilon > 0$, and may take $\varepsilon < 1$. Then $B_\varepsilon(x; d_2) = B(x; d) \subset U$, so U is d_2-open.

Since $d_3(x, y) \leqslant d(x, y)$, U d_3-open $\Rightarrow$ U d-open as for d_2. Suppose U d-open and $x \in U$. Then $B_\varepsilon(x; d) \subset U$ for some $\varepsilon > 0$. Let $\delta = \mathrm{Min}\{\frac{1}{2}\varepsilon, \frac{1}{2}\}$. Then $d_3(x, y) < \delta \Rightarrow d_3(x, y) < \frac{1}{2}$, so

$$d(x, y) = d_3(x, y)/(1 - d_3(x, y)) \leqslant 2d_3(x, y) < \varepsilon.$$

Thus $B_\delta(x; d_3) \subset B_\varepsilon(x; d) \subset U$, and U is d_3-open. ($\delta = \varepsilon/(1+\varepsilon)$ would do equally well: since $a/(1+a)$ is strictly increasing for $a \geqslant 0$, $d_3(x, y) < \varepsilon/(1+\varepsilon) \Leftrightarrow d(x, y) < \varepsilon$.)

21. To economise on algebra, note $\rho(x) = xA$ where

$$A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

and $\{d_2(x, y)\}^2 = (x - y)(x - y)^t$, where t denotes transpose.

22. Both translation-invariant, neither rotation-invariant: can be seen by picture or by analysis.

23. (M3) If $m - n = p^{r-1}k$, $n - q = p^{s-1}k'$, then $m - q = p^{t-1}k''$, where $t = \mathrm{Min}\{r, s\}$ $(p \nmid k, k', k'')$. So

$$d(m, q) = \mathrm{Max}\{d(m, n), d(n, q)\} \leqslant d(m, n) + d(n, q).$$

24. Let

$$f(x, y) = \begin{cases} x^2/y, & y \neq 0, \\ 0, & y = 0. \end{cases}$$

Then f restricted to any straight line through 0 in $\mathbf{R}^2$ is continuous, but $f(0, 0) = 0$, while in any neighbourhood of $(0, 0)$ we can find a point (k, k^2) for k sufficiently small, and $f(k, k^2) = 1$.

25. Not necessarily; iff $g(x) - f(x)$ constant.

26. Not necessarily. In discrete M, $B_r(x) = B_s(y) = M$ for all x, y in M and all $r, s > 1$.

27. The bumps are ±1 mm.
Yes (if you have more than £30 000 to spare).

$$I \equiv \int_0^1 \{1 + (g'(x))^2\}^{\frac{1}{2}}\, dx = \int_0^1 \{1 + 10^{20}\cos^2(10^{16}x)\}^{\frac{1}{2}}\, dx$$

$$\geqslant \int_0^1 10^{10} |\cos 10^{16} x|\, dx,$$

and $|\cos 10^{16}x| \geqslant \frac{1}{2}$ for two-thirds of the range $[0, 1]$, so $I \geqslant 10^{10}/3$, costing $> 3 \cdot 10^6$ p $=$ £30 000.

Exercises 3.9

2. (ii) 29 in number.
8. Given U open in $\mathbf{R}^2$ and $x = (x_1, x_2)$ in U, then $B_\varepsilon(x) \subset U$ for some $\varepsilon > 0$. Choose rational q_1, q_2 within $\varepsilon/5$ of x_1, x_2 respectively, and put $s = (q_1, q_2)$. Then $d_2(x, s) < (\varepsilon\sqrt{2})/5 < \varepsilon/3$. Choose rational q with $\varepsilon/3 < q < 2\varepsilon/3$. Then $x \in B_q(s) \subset B_\varepsilon(x) \subset U$. So U is a union of sets from the given collection.
11. Let $(\mathscr{T}_K)_H$ be the topology induced on H by $\mathscr{T}_K$. Then $V \subset H$ is in $(\mathscr{T}_K)_H$ iff $V = H \cap U$ for U in $\mathscr{T}_K$, and $U \in \mathscr{T}_K$ iff $U = K \cap W$ for W in $\mathscr{T}$. Hence $V \in (\mathscr{T}_K)_H$ iff $V = H \cap K \cap W = H \cap W$ for W in $\mathscr{T}$, i.e. iff $V \in \mathscr{T}_H$.
12. First, $H \cap B \in \mathscr{T}_H$ for any B in $\mathscr{B}$, since $\mathscr{B} \subset \mathscr{T}$. For any V in $\mathscr{T}_H$, $V = H \cap U$ for U in $\mathscr{T}$, and $U = \bigcup_I B_i$ for some indexing set I, where each $B_i \in \mathscr{B}$. Then $V = H \cap U = \bigcup_I (H \cap B_i)$.
13. If U in $\mathscr{T}_\mathrm{H}$, then $U = V \cap H$, some V in $\mathscr{T}$. For $x \in U$, $B_\varepsilon(x; d) \subset V$ for some $\varepsilon > 0$, so $B_\varepsilon(x; d_\mathrm{H}) \subset U$ and U is d_H-open. If U d_H-open, then for each x in U, $B_{\varepsilon(x)}(x; d_\mathrm{H}) \subset U$ for some $\varepsilon(x) > 0$ and $U = H \cap V$ where $V = \bigcup_{x \in U} B_{\varepsilon(x)}(x; d)$ is in $\mathscr{T}$.
14. Since these metrics equivalent, enough to check for d_∞. For this, enough to show that each set in a basis for the one topology is open in the other topology and vice-versa. Now
$$B_\varepsilon((x, y); d_\infty) = B_\varepsilon(x; d) \times B_\varepsilon(y; d') \qquad \text{(check)},$$
so every $B_\varepsilon((x, y); d_\infty)$ is open in product topology. Conversely, every $U \times V$ in defining basis for product topology is d_∞-open, since given $(x, y) \in U \times V$, there exist ε, $\varepsilon' > 0$ such that $B_\varepsilon(x; d) \subset U$, $B_{\varepsilon'}(y; d) \subset V$. Let $\delta = \operatorname{Min}\{\varepsilon, \varepsilon'\}$. Then
$$B_\delta((x, y); d_\infty) = B_\delta(x; d) \times B_\delta(y; d') \subset U \times V.$$
21. Geometrically clear, awkward to write a proof. Sketch of one method: first stretch each square radially out to corresponding disc (see Fig. 12.2). Then do a radial deformation which opens out the sector S_{BXA} subtended by the arc BXA to S_{BAD} and squashes S_{BCA} to S_{BCD}.

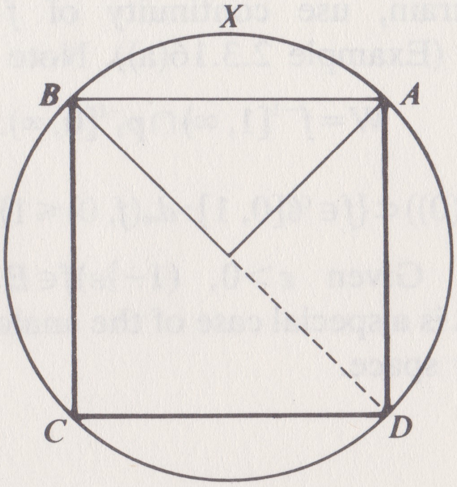


FIG. 12.2

24. $H = (0, 2) \cup (3, 4)$, $K = (1, 3)$.

25. If f continuous, and $y \in f(Cl(H))$, then $y = f(x)$ for x in $Cl(H)$. If U open and $y \in U$, then $f^{-1}(U)$ open and x in it, so $f^{-1}(U) \cap H \neq \emptyset$ since $x \in Cl(H)$. Hence

$$U \cap f(H) \supset f(f^{-1}U \cap H) \neq \emptyset,$$

and $y \in Cl(f(H))$. Conversely if $f(Cl(H)) \subset Cl(f(H))$ for all $H \subset T_1$, and if V closed in T_2, then $f(Cl(f^{-1}V)) \subset Cl(V) = V$, so $Cl(f^{-1}V) \subset f^{-1}V$, i.e. $f^{-1}V$ closed. Hence by 3.7.8 f continuous.

26. $f(Cl(H)) \subset Cl(f(H))$; $f(Cl(H))$ closed and contains $f(H)$ so contains $Cl(f(H))$; so $f(Cl(H)) = Cl(f(H))$.

27. False, false, true.

30. (i) First $T - Cl(T-H) \subset T-(T-H) = H$. Since

$$Cl(H) = \{Cl(H) \cap Cl(T-H)\} \cup \{Cl(H) \cap (T - Cl(T-H))\},$$

it follows that $Cl(H) \subset H$ iff $Cl(H) \cap Cl(T-H) \subset H$, i.e. iff $b(H) \subset H$.

33. (i) $f^{-1}(W) = \bigcup_{i=1}^{n} (f \mid V_i)^{-1}(W)$ (check). If W closed in T_2 $(f \mid V_i)^{-1}(W)$ closed in V_i, hence in T_1 since V_i closed in T_1.
(ii) Similar, using unlimited union and open W.

34. (*b*) *Let* $W = \{(x, y) \in \mathbf{R}^2 : x > 0, y \geqslant 1/x\}$. Then $p_1(W) = (0, \infty)$. From picture, W clearly closed in $\mathbf{R}^2$. To prove this

without strain, use continuity of $f:\mathbf{R}^2\to\mathbf{R}$ given by $f(x, y)=xy$ (Example 2.3.16(a)). Note

$$W=f^{-1}[1,\infty)\cap p_1^{-1}[0,\infty).$$

36. As in 3.7.23,

$$Cl(B_1(0))\subset\{f\in\mathscr{C}[0,1]:d_\infty(f,0)\leqslant 1\}=S,\quad\text{say.}$$

Suppose $f\in S$. Given $\varepsilon>0$, $(1-\frac{1}{2}\varepsilon)f\in B_1(0)\cap B_\varepsilon(f)$ so $f\in Cl(B_1(0))$. This is a special case of the analogous result for any normed vector space.

Exercises 4.3

5. (b) If $z\in K$ then $d(y,z)\geqslant d(x,z)-d(x,y)$, so $d(y,z)\geqslant d(x,K)-d(x,y)$ for any z in K, hence $d(y,K)\geqslant d(x,K)-d(x,y)$.

(c) $|f_K(x)-f_K(y)|=|d(x,K)-d(y,K)|\leqslant d(x,y)$ from (b) and (b) with x, y interchanged.

From (c) applied to H and K, g is continuous. $g^{-1}(0,\infty)$ and $g^{-1}(-\infty,0)$ are disjoint. If $x\in H$ then $f_H(x)=0$ and since $H\cap K=\varnothing$, $x\notin K$ so since K is closed, by (a) $f_K(x)>0$. Hence $g(x)>0$, i.e. $H\subset g^{-1}(0,\infty)$. Similarly $K\subset g^{-1}(-\infty,0)$.

6. Prove complement open: if $x\in T_1-W$ then $f(x)\neq g(x)$ so there exist disjoint open U, V with $f(x)\in U$, $g(x)\in V$. Then $f^{-1}(U)\cap g^{-1}(V)$ open, contains x, contained in T_1-W.

Exercises 5.10

4. Identity map is $(\mathscr{T}_2,\mathscr{T}_1)$-continuous and onto.

5. (iv) This is $f^{-1}(0)$ where $f(x,y)=x^2+y^2-1$ defines $f:\mathbf{R}^2\to\mathbf{R}$, so $f^{-1}(0)$ closed since f continuous. Also bounded, so compact.

(v) similar, (vi) not closed, (vii) not bounded.

7. Follows from Exercise 6.

8. $C\cap C'$ closed in T hence in C, which is compact.

11. A not bounded: f the inclusion.
A not closed: $f(x)=1/(x-b)$ where $b\in Cl(A)-A$.

13. (b) The open cover $\{(1,2,\ldots,n):n\in\mathbf{N}\}$ has no finite subcover.

(c) If $f:T\to T'$ continuous where T' Hausdorff, and if $f(x)\neq f(y)$ then deduce $x\in f^{-1}(U)$, $y\in f^{-1}(V)$ and these are disjoint open sets. But no such sets in T.

14. $g=d\circ(f\times 1)\circ\Delta$ where $\Delta: T\to T\times T$ given by $\Delta(x)=(x, x)$. Hence (check) g continuous, so attains lower bound ε, so $\varepsilon>0$.

15. Two possible formulations for metric space M. If policemen uniformly near-sighted: given $\varepsilon>0$, there exists a finite set $S\subset M$ such that $M=\bigcup_{x\in S} B_\varepsilon(x)$. This will not do as definition of compactness; e.g. $\mathbf{Q}\cap[0, 1]$ satisfies it. If policemen have different lengths of vision: given any open cover $\{B_{\delta(x)}(x): x\in M\}$ there is a finite subcover. This will do as definition of compactness for metric spaces (check).

17. Suppose W closed in T, $x\in T-W$. For each $y\in W$, let $U(y)$, $V(y)$ be disjoint open sets with $y\in U(y)$, $x\in V(y)$. Let $\{U(y_1), U(y_2),\ldots, U(y_n)\}$ be a subcover of $\{U(y)\}$ for W. Put $U=\bigcup_{i=1}^{n} U(y_i)$, $V=\bigcap_{i=1}^{n} V(y_i)$. Then U, V disjoint open, contain W, x respectively. Similarly for normality.

19. Negation: for some $\varepsilon_0>0$ and all $\delta>0$, there exist x, y in D with $|x-y|<\delta$, $|f(x)-f(y)|\geqslant\varepsilon_0$. Take $\varepsilon_0=1$, $\delta>0$. Choose $n>1/\delta$. Then $|f(1/2n)-f(1/n)|=n\geqslant 1$, $|1/2n-1/n|<\delta$.

21. (b) No; $\sin 1/x$.

22. (a), (d) uniformly continuous.

24. Solution to Exercise 2.6.20 does first part. On $\mathbf{R}$, $d_3(x, y)/d(x, y)=1/(1+d(x, y))$ has greatest lower bound 0, so d_3 not Lipschitz equivalent to d.

25. Identity map not uniformly (d_2, d)-continuous, where d_2 is usual metric, since take $\varepsilon=1$, say. Not possible to get $\delta>0$ such that $|x-y|<\delta\Rightarrow d(x, y)<1$ – see answer to Exercise 19.

    Since $|x-y|\leqslant d(x, y)$, U d_2-open $\Rightarrow$ U d-open. Given a in $(0, 1)$, $\varepsilon>0$, put $\delta=\text{Min}\{\tfrac{1}{2}a^2\varepsilon, 1/2a\}$. If $|x-a|<\delta$, then $d(x, a)=|x-a|/|xa|<2\,|x-a|/a^2<\varepsilon$. Thus identity map (d_2, d)-continuous, i.e. U d-open $\Rightarrow$ U d_2-open.

27. With notation of Exercise 4.3.5, $f_K\,|\,H: H\to\mathbf{R}$ is continuous and H compact, so $f_K\,|\,H$ attains its lower bound (which is $d(H, K)$) on H, $d(H, K)=d(x, K)$ for some x in H. Since K closed, $d(x, K)=0$ iff $x\in K$. Hence $d(H, K)=0$ iff $H\cap K\neq\varnothing$.

    No. Let $H\subset\mathbf{R}^2$ be $\{(x, y): x\leqslant 0\}$,

$$K=\{(x, y): x>0,\ y\geqslant 1/x\}.$$

Then $H\cap K=\emptyset$. But given $\varepsilon>0$, let $y>1/\varepsilon$. Then $d_2((0, y), (1/y, y))=1/y<\varepsilon$, so $d(H, K)=0$.

28. Prove $f(A)=A$ and use Exercise 7 to show $A\neq\emptyset$.

29. $\{f^{-1}(a, \infty): a\in \mathbf{R}\}$ is an open cover of T. Pick a finite subcover and let a_0 be least a involved. Then $T=f^{-1}(a_0, \infty)$, so $f(T)$ bounded below by a_0. (Could replace original cover by $\{f^{-1}(a, \infty): a\in A\}$ for any set A not bounded below.)

    Let $a_0=\inf f(T)$. If a_0 not attained then $\{f^{-1}(a_0+\varepsilon, \infty): \varepsilon>0\}$ is an open cover of T. Let ε_0 be least ε involved in a finite subcover. Then $a_0+\varepsilon_0$ is a lower bound for $f(T)$, contradicting choice of a_0.

30. Given $x\notin p_1(W)$, for each y in T_2 then $(x, y)\notin W$, so there exist open $U(y)$, $V(y)$ such that

$$(x, y)\in U(y)\times V(y)\subset T_1\times T_2-W.$$

Let $\{V(y_1), V(y_2), \ldots, V(y_n)\}$ be a subcover of $\{V(y)\}$ for T_2, and put $U=\bigcap_{i=1}^{n} U(y_i)$. Then

$$U\times T_2\subset T_1\times T_2-W,$$

so $x\in U\subset T_1-p_1(W)$. Hence $T_1-p_1(W)$ open.

32. Let $e_i\in\mathbf{R}^n$ have ith coordinate 1, rest 0. For arbitrary norm $\|\ \|$ on $\mathbf{R}^n$ let $k=\operatorname{Max}_i \|e_i\|$. Then for $x=(x_1, x_2, \ldots, x_n)$,

$$\|x\|=\|\sum x_ie_i\|\leqslant\sum |x_i|\ \|e_i\|\leqslant k\ \|x\|_1,$$

where $\|x\|_1=|x_1|+|x_2|+\ldots+|x_n|$. Prove $\|\ \|:\mathbf{R}^n\to\mathbf{R}$ is continuous with respect to 1-norms. The restriction of this continuous function to compact $S=\{x\in\mathbf{R}^n: \|x\|_1=1\}$ therefore attains lower bound m, so $m>0$. Hence $\|x\|\geqslant m$ for all x in S. Now use linearity to prove $\|x\|\geqslant m\ \|x\|_1$ for all x in $\mathbf{R}^n$. Hence arbitrary norm Lipschitz equivalent to 1-norm, so any two norms on $\mathbf{R}^n$ are Lipschitz equivalent.

34. The numeral 6 suggests such a map.

35. (T1) $A\cup\{\infty\}\in\mathcal{T}'$ since $A-A=\emptyset$ is compact and closed in T. $\emptyset\in\mathcal{T}'$ since $\emptyset\in\mathcal{T}$.

    (T2) Suppose U, $U'\in\mathcal{T}'$. Three cases.

    (a) If $\infty\in U\cap U'$, then $A-U$, $A-U'$ both compact and closed in T, so $A-(U\cap U')=(A-U)\cup(A-U')$ compact closed in T, so $U\cap U'\in\mathcal{T}'$.

(b) If $\infty\in U$, $\infty\notin U'$, then $A-U$ closed in T so $U\cap A$ open in T, as is U', hence $U\cap U'\subset A$ and open in T, so in $\mathcal{T}'$. (If $\infty\notin U$, $\infty\in U'$, just interchange names.)

(c) If $\infty\notin U$, $\infty\notin U'$, then U, $U'\in\mathcal{T}$ so $U\cap U'\in\mathcal{T}\subset\mathcal{T}'$.

(T3) Suppose $U_i\in\mathcal{T}'$ for each i in some indexing set I. Two cases.

(a) If for all i, $\infty\notin U_i$, then $U_i\in\mathcal{T}$ for all i, so $\bigcup_I U_i\in\mathcal{T}\subset\mathcal{T}'$.

(b) If $\infty\in U_j$ for some j in I, then $A-U_j$ compact and closed in T. Then $A-\bigcup_I U_i=\bigcap_I (A-U_i)$ is closed in T and also in compact $A-U_j$ so is compact, so $\bigcup_I U_i\in\mathcal{T}'$.

Thus T' is a topological space, and T is a subspace (check). Given any open cover $\mathcal{U}$ of T', then $\infty\in U_0$ for some $U_0\in\mathcal{U}$, so $A-U_0$ is compact and covered by $\mathcal{U}$, so there exists a finite subcover $\{U_1, U_2,\ldots, U_n\}$ of $\mathcal{U}$ for $A-U_0$, and now $\{U_0, U_1,\ldots, U_n\}$ is a finite subcover of $\mathcal{U}$ for T'. Thus T' is compact.

36. Suppose T is locally compact and Hausdorff. Given x, $y\in T'$, with $x\neq y$, then either x, $y\in T$, and there are disjoint U, V open in T containing x, y, and U, $V\in\mathcal{T}\subset\mathcal{T}'$, or say $x=\infty$, $y\in T$. Then by local compactness, there is an open $U\subset T$ with $y\in U$ and $Cl(U)$ compact. Then $V=(T-Cl(U))\cup\{\infty\}$ is open in T' with $\infty\in V$, $U\cap V=\varnothing$ as required. (In fact the converse is true too: if T' is Hausdorff then T is Hausdorff and locally compact.)

Exercises 6.6

3. (v) Path-connected, hence connected. Think of it as a fine grid of horizontal and vertical lines.

5. The identity map is $(\mathcal{T}_2, \mathcal{T}_1)$-continuous and onto, so $\{A, \mathcal{T}_2\}$ connected $\Rightarrow$ $\{A, \mathcal{T}_1\}$ connected. Converse false in general: e.g. $\mathbf{R}$ with usual and discrete topologies.

6. Let $b=f(a)$ for some a in T, and let

$$U=\{x\in T: f(x)=b\}.$$

Show U and $T-U$ open since f locally constant.

7. If T finite, $\{x\} \mid T-\{x\}$ partitions T since $T-\{x\}$ a finite union of closed sets so closed.

11. (a) If f takes two real values $a<b$, then by I.V.P. it takes on all values in $[a, b]$, which is uncountably infinite, hence so is T.
    (b) Apply (a) to $f: M \to \mathbf{R}$ given by $f(x)=d(x, a)$ for fixed a in M.

12. Consider e.g. any non-degenerate triangle ABC in $\mathbf{R}^2$. Then BC and $BA \cup AC$ both connected, so both $f(BC)$ and $f(BA \cup AC)$ contain $[f(B), f(C)]$ by I.V.P.

14. After some experimentation with pictures, following suggested. (Draw.) Suppose that f is continuous and has the stated property. For some h in $\mathbf{R}$ let $f(a)=f(b)=h$, with $a<b$. Then $f \mid [a, b]$ attains its bounds on $[a, b]$. It must have $\max = k > h$ or its $\min = m < h$, otherwise f constant on $[a, b]$ and contradiction. Suppose e.g. $k>h$, and $f(c)=k$ where $c \in (a, b)$. Now k must be attained once more, say $f(c')=k$. Suppose first $c'<a$. Choose any value y between h and k. Then y attained once in each of $[c', a]$, $[a, c]$, $[c, b]$, by I.V.P., contradiction. Similar argument if $c'>b$. Suppose $c' \in (a, c)$. Then for any d in (c', c), have $f(d)<k$, otherwise contradiction. Let y be any value between $\operatorname{Max}\{h, f(d)\}$ and k. Then y attained once in each of $[a, c]$, $[c, d]$, $[d, c']$, $[c', b]$, contradiction. Similar argument if $c' \in (c, b)$. This completes contradiction if $k>h$. Similar argument if $m<h$.

15. Suppose f such a function. If f constant on $\mathbf{R}-\mathbf{Q}$, then constant on $\mathbf{R}$ by continuity. So let $f(x)=q$, $f(y)=r$, where $q \neq r$, $q, r \in \mathbf{Q}$. There exist uncountably many irrationals between q and r, and by I.V.P. these must all be attained. But only countably many rationals to attain them at, so impossible.

16. Define $h:[0, 1] \to \mathbf{S}$ by $h(t)=(\cos 2\pi t, \sin 2\pi t)$. Then h continuous by Propositions 3.4.2(b), 3.5.4, since cos and sin are continuous. Also, h onto, so $\mathbf{S}$ connected. Since $g(-x)=-g(x)$, g takes value 0 by I.V.P. (in the general form of Corollary 6.2.13).

20. First note $H \cap K \neq \varnothing$, otherwise $H \mid K$ partitions $H \cup K$. Suppose $A \mid B$ partitions H. Then A, B closed in H hence in T since H closed in T. May assume $A \cap H \cap K \neq \varnothing$, otherwise

interchange names of A, B. Then $A \supset H \cap K$ otherwise $A \cap H \cap K \mid B \cap H \cap K$ partitions $H \cap K$. So $B \cap K = \emptyset$ since $B \subset H$ and $B \cap A = \emptyset$. Thus $B \cap (K \cup A) = \emptyset$. Then $(K \cup A) \mid B$ partitions $H \cup K$ (K, A closed in T so $K \cup A$ closed in T, hence in $H \cup K$; also B closed in T hence in $H \cup K$.) Similar contradiction if K not connected. Shorter solution using 6.2.1 and 3.9.33(i).

21. $C_n = \{(x, y) \in \mathbf{R}^2$: either $x > 0, y \geqslant 1/x$ or $y \leqslant 1/n\}$ (Draw.)
22. Suppose $A \mid B$ partitions $\bigcap C_n$. Since latter closed, A, B closed in C_1. Since C_1 compact Hausdorff so normal (Exercise 5.10.17), there are disjoint U, U' open in C_1 with $A \subset U$, $B \subset U'$. Let $V_n = C_n - (U \cup U')$. This is closed since $U \cup U'$ open in C_1 and $C_n \subset C_1$. Also, $V_n \neq \emptyset$ since otherwise $U \cap C_n \mid U' \cap C_n$ would partition C_n (check). Since $V_n \supset V_{n+1}$ for all n, Exercise 5.10.7 says $\bigcap V_n \neq \emptyset$. But $\bigcap V_n = (\bigcap C_n) - (U \cup U') = \emptyset$.

Exercises 7.3

4. Sufficiency: suppose f not continuous at a. Then for some $\varepsilon_0 > 0$, and any n, there exists a point, say x_n, in M_1 such that $d_1(x_n, a) < 1/n$ but $d_2(f(x_n), f(a)) \geqslant \varepsilon_0$. Then (x_n) converges to a but $(f(x_n))$ does not converge to $f(a)$.
10. Problem mainly notational. Suppose (x^r) a sequence in C, where $x^r = (x_1^r, x_2^r, \ldots, x_n^r, \ldots)$ so $|x_n^r| \leqslant 1/n$ for all, r, n. Since x_1^1, $x_1^2, \ldots$ is a sequence in $[-1, 1]$, it has a subsequence, say $x_1^{r(1,1)}$, $x_1^{r(1,2)}, \ldots$ converging to a point x_1 in $[-1, 1]$. Inductively suppose sequences of integers have been chosen as follows:

$$\begin{array}{lll} r(1, 1), & r(1, 2), & r(1, 3), \ldots \\ r(2, 1), & r(2, 2), & r(2, 3), \ldots \\ \vdots & \vdots & \vdots \\ r(n-1, 1), & r(n-1, 2), & r(n-1, 3), \ldots. \end{array}$$

such that each is a subsequence of the sequence above it, and for $1 \leqslant i < n$, $x_i^{r(i,1)}, x_i^{r(i,2)}, \ldots$ converges to x_i in $[-1/i, 1/i]$. Then $x_n^{r(n-1,1)}$, $x_n^{r(n-1,2)}, \ldots$ is a sequence in $[-1/n, 1/n]$ so has a subsequence, say $x_n^{r(n,1)}$, $x_n^{r(n,2)}, \ldots$ converging to a point x_n in $[-1/n, 1/n]$. Now consider the diagonal subsequence $(x^{r(n,n)})$ of (x^r). This converges to $x = (x_1, x_2, \ldots)$ in C. For given $\varepsilon > 0$, there exists N such that $\sum_{N+1}^{\infty} 4/n^2 < \frac{1}{2}\varepsilon^2$ $\left(\text{since} \sum_{1}^{\infty} 4/n^2 \text{ converges}\right)$.

For each $i \leqslant N$, $(x_i^{r(n,n)})_{n \in \mathbf{N}}$ is a real sequence which beyond stage $n = i$ is a subsequence of $(x_i^{r(i,n)})$ hence converges to x_i. So there exists N_i such that $|x_i^{r(n,n)} - x_i| < (2N)^{-\frac{1}{2}}\varepsilon$ for $n \geqslant N_i$. Then for $n \geqslant \text{Max}\{N, N_1, N_2, \ldots, N_N\}$,

$$d_2(x^{r(n,n)}, x) = \left\{\sum_{i=1}^{\infty} (x_i^{r(n,n)} - x_i)^2\right\}^{\frac{1}{2}}$$

$$\leqslant \left\{\sum_{i=1}^{N} (x_i^{r(n,n)} - x_i)^2 + \tfrac{1}{2}\varepsilon^2\right\}^{\frac{1}{2}} < \varepsilon.$$

11. Suppose $x, y \in \bigcap V_n$. Then $x, y \in V_n$ for any n, so $d(x, y) \leqslant$ diam V_n. This is true for all n, so

$$d(x, y) \leqslant \inf\{\text{diam } V_n : n \in \mathbf{N}\} = \alpha, \quad \text{say.}$$

This is true for all x, y in $\bigcap V_n$ so $\beta = \text{diam} \bigcap V_n \leqslant \alpha$. Suppose $\beta < \alpha$, say $\beta = \alpha - \varepsilon$. Then diam $V_n \geqslant \beta + \varepsilon$ for all n, so for each n there exist x_n, y_n in V_n with $d(x_n, y_n) > \beta + \frac{1}{2}\varepsilon$. The sequence (x_n) has a convergent subsequence $(x_{n(r)})$ converging to x, say, and $(y_{n(r)})$ has a subsequence $(y_{m(r)})$, converging to y say, since M is compact. Hence $(x_{m(r)})$ also converges to x. Check $d(x_{m(r)}, y_{m(r)})$ converges to $d(x, y)$. Hence $d(x, y) \geqslant \beta + \frac{1}{2}\varepsilon$. But since each V_n closed and $x_{m(r)} \in V_n$ for all r sufficiently large, $x \in \bigcap V_n$. Similarly $y \in \bigcap V_n$. Contradiction, since $d(x, y) \geqslant \beta + \frac{1}{2}\varepsilon$, diam $\bigcap V_n = \beta$.

12. (a) Given any x in M, consider $x, f(x), f^{(2)}(x) = f \circ f(x), \ldots$. Since M compact, this has a convergent subsequence, say $f^{(n_1)}(x)$, $f^{(n_2)}(x), \ldots, f^{(n_r)}(x), \ldots$. This is a Cauchy sequence, so given $\varepsilon > 0$, there exists R such that $d(f^{(n_r)}(x), f^{(n_s)}(x)) < \varepsilon$ for all $r > s \geqslant R$, so $d(f^{(n_r - n_s)}(x), x) < \varepsilon$. So for any $\varepsilon > 0$, $B_\varepsilon(x)$ contains a point in $f(M)$. Hence $x \in Cl(f(M))$. But M compact Hausdorff, so $f(M)$ compact and hence closed. So $x \in f(M)$ for any x in M, i.e. f onto.
    (b) Apply (a) to prove $g \circ f$ and $f \circ g$ both onto, hence g, f onto.
    (c) Let $f(x) = x + 1$.

Exercises 8.6

2. (a) For some N, $|f_n(x) - f(x)| < 1$ for all $n \geqslant N$, all x in D. Let $|f_N(x)| \leqslant \Delta$ for all x in D. Then

$$|f(x)| \leqslant 1 + |f_N(x)| \leqslant 1 + \Delta$$

for all x in D.

(b) This is just the statement that the set of members of a convergent sequence in a metric space is bounded.

3. The pointwise limits are the zero functions except for (v) where the limit ($f(x)=0,\ x\in[0, 1),\ f(1)=\frac{1}{2}$) is discontinuous, hence convergence non-uniform. Also non-uniform in (ii), (iv), otherwise uniform. For (vi) use Exercise 1: on $[0, \frac{1}{2}]$, $|n^{-x}x^n \cos nx| \leqslant (\frac{1}{2})^n$; on $[\frac{1}{2}, 1]$, $|n^{-x}x^n \cos nx| \leqslant n^{-\frac{1}{2}}$.

6. $h_n \to h$ pointwise, where

$$h(x)=\begin{cases} 0 & \text{if} \quad x=0 \quad \text{or} \quad x\notin \mathbf{Q}, \\ p & \text{if} \quad x=p/q,\ (p, q)=1. \end{cases}$$

But $|h_n(p/q)-h(p/q)| \geqslant p/n$. In any interval $[a, b]$ with $b>a$, there exist rationals p/q (in lowest terms) with $p \geqslant n$. Hence no N such that $|h_n(x)-h(x)|<1$ for all x in $[a, b]$ and all $n \geqslant N$.

7. $$f_n(x)=\begin{cases} 1/n, & x\in \mathbf{Q}, \\ 0, & x\notin \mathbf{Q}. \end{cases}$$

9. Put $g_n(x)=f_n(x)-f(x)$, so g_n continuous on T, $g_n(x) \geqslant g_{n+1}(x)$ for all n and x, and (g_n) converges pointwise to zero. Plan: use monotonicity and continuity to show (g_n) locally uniformly convergent, and then use compactness of T. For any x in T, given $\varepsilon>0$, there exists $N(x)$ such that $0 \leqslant g_n(x)<\frac{1}{2}\varepsilon$ for all $n \geqslant N(x)$. Since $g_{N(x)}$ continuous at x, there exists open $U(x)$ with $x\in U(x)$ and $0 \leqslant g_{N(x)}(y)<\varepsilon$ for all y in $U(x)$, and then by monotonicity, $0 \leqslant g_n(y)<\varepsilon$ for all $n \geqslant N(x)$ and all y in $U(x)$. This is what 'locally uniformly convergent' means. Now let $\{U(x_1), U(x_2), \ldots U(x_r)\}$ be a finite subcover of $\{U(x): x\in T\}$ for T, let

$$N=\operatorname{Max}\{N(x_1), N(x_2), \ldots, N(x_r)\}$$

and get $0 \leqslant g_n(y)<\varepsilon$ for all $n \geqslant N$ and all y in T.

Exercises 9.6

1. (a) If (x_n) a Cauchy sequence in $M\cup M'$, then either $x_n\in M$ for infinitely many n or likewise for M', or both. In any case this gives a convergent subsequence by completeness of M and M'; now use Lemma 9.2.5.

(b)) $\bigcap_i M_i$ is a closed subspace of complete M_1, say.

3. (i) and (iii) complete ((iii) as closed subspace of complete $\mathbf{R}^2$).

5. Usual metric d_2 and $d(x, y)=|x^{-1}-y^{-1}|$ on $(0, 1)$. Then $(1/n)$ is d_2-Cauchy but not d-Cauchy.
6. Define $f:[1, \infty)\to(0, 1]$ by $f(x)=1/x$.
8. Follow proof (c) for Example 9.2.12, using C-B-S inequality to check triangle inequality in $\mathcal{R}$ for L^2 pseudo-metric.
9. Suppose $x=(x_n)\in Cl(C)$. Given $\varepsilon>0$, there exists (y_n) in C with $\sup\limits_n |x_n-y_n|<\varepsilon/3$. Since $(y_n)\in C$, (y_n) is Cauchy, so there exists N such that $|y_n-y_m|<\varepsilon/3$ for $m, n\geq N$. Then $|x_n-x_m|<\varepsilon$ for all $m, n\geq N$, so (x_n) Cauchy hence convergent since **R** complete.
10. F not closed in ℓ^2. For if $x^n=(1, \frac{1}{2}, \dots, 1/n, 0, 0, 0, \dots)$ then (x^n) converges to $(1, \frac{1}{2}, \dots, 1/n, 1/(n+1), \dots)$.
12. First, $f(0, \frac{1}{4})\subset(0, \frac{1}{4})$. Also, $f'(x)=2x<\frac{1}{2}$ for all x in $(0, \frac{1}{4})$. By mean value theorem (M.V.T.) given x, y in $(0, \frac{1}{4})$, $|f(x)-f(y)|=|x-y|\,|f'(\xi)|$ for some ξ in $(0, \frac{1}{4})$ so $|f(x)-f(y)|<\frac{1}{2}|x-y|$ for all x, y in $(0, \frac{1}{4})$.
14. Since $f'(x)=\frac{1}{4}x^{-\frac{1}{2}}(2+x^{\frac{1}{2}})^{-\frac{1}{2}}$, $0\leq f'(x)\leq\frac{1}{4}$ if $x\geq1$. Also, $f(x)\geq1$ if $x\geq1$, so f a contraction on $[1, \infty)$, hence has a unique fixed point in $[1, \infty)$, to which the sequence given by $x_n=f(x_{n-1})$ converges for any $x_1\geq1$ (in fact for any $x_1\geq0$, since then $x_2\geq1$). The root $x=1$ of $x^4-4x^2-x+4=0$ is not a root of $x=(2+x^{\frac{1}{2}})^{\frac{1}{2}}$, so causes no problems here.
15. By contraction theorem, there is a unique p in M such that $f^{(r)}(p)=p$. Then $f^{(r)}(f(p))=f(f^{(r)}(p))=f(p)$, so by uniqueness of fixed point of $f^{(r)}$, $f(p)=p$. Any fixed point of f is also a fixed point of $f^{(r)}$, so only one fixed point of f.

    Let K be a contraction constant for $f^{(r)}$. Given any x in M, let $\Delta=\max\limits_{0<i<r} d(f^{(i)}(x), x)$. Let N be such that $d(f^{(rn)}(x), p)<\frac{1}{2}\varepsilon$ for all $n\geq N$, and also $K^N\Delta<\frac{1}{2}\varepsilon$. Then for $m\geq rN$, we have $m=rn+i$ for some $n\geq N$ and $0\leq i<r$, so

    $$\begin{aligned} d(f^{(m)}(x), p) &\leq d(f^{(m)}(x), f^{(rn)}(x))+d(f^{(rn)}(x), p) \\ &< K^n d(f^{(i)}(x), x)+\tfrac{1}{2}\varepsilon \\ &\leq K^N\Delta+\tfrac{1}{2}\varepsilon \\ &< \varepsilon. \end{aligned}$$

    If willing to assume f continuous, then alternative proof available for last part.

16. Given any $K<1$ a candidate for contraction constant, choose δ such that $\sin x \geqslant \frac{1}{2}(1+K)$ for all x in $[\frac{1}{2}\pi-\delta, \frac{1}{2}\pi]$. Then

$$|\cos \tfrac{1}{2}\pi - \cos(\tfrac{1}{2}\pi-\delta)| > K\delta$$

by M.V.T., hence cos not a contraction.

Put $g(x)=\cos(\cos x)$. Then $g'(x)=\sin x.\sin(\cos x)$, and $|g'(x)|\leqslant \sin 1<1$. Hence by M.V.T., g a contraction.

Even starting with a reasonable guess such as $\frac{1}{4}\pi$, iteration of cos is slow work.

20. $\mathbf{Q}$, $(\mathbf{R}-\mathbf{Q})$.

22. If U_n open and $\mathbf{Q}\subset U_n$ then $Cl(\mathbf{R}-U_n)=\mathbf{R}-U_n$ contains no rationals, so is nowhere dense (check). Then $\mathbf{R}$ would be the countable union of nowhere dense sets $(\mathbf{R}-U_n)\cup\{q_n\}$.

Exercises 10.3

3. For each n, let S_n be a finite $1/n$-net. Then $\bigcup_{n=1}^{\infty} S_n$ is a countable dense set; now use Exercise 3.9.37.

5. There is no finite 1-net, since the points $(0, 0, \ldots 0, 1, 0, \ldots)$ are all at distance 1 from one another.

7. (i) Given $\varepsilon>0$, take $\delta=\text{Min}\{\varepsilon, 1\}$ and for any x in $[0, \delta)$, any n in $\mathbf{N}$, $|f_n(x)-f_n(0)|=x^n<\delta^n<\delta=\varepsilon$. Hence equicontinuous at 0.

   Take $\varepsilon=\frac{1}{2}$ and any $\delta>0$. Can pick x in $(1-\delta, 1)$ and then suitably large n so that

$$|f_n(x)-f_n(1)|=1-x^n\geqslant\tfrac{1}{2},$$

   so not equicontinuous at 1.

   (ii) Let $a\in(0, 1)$. Since $f'_n(x)=nx^{n-1}$, for all x in $[0, \frac{1}{2}(a+1)]$ we have $f'_n(x)\leqslant n(\frac{1}{2}(a+1))^{n-1}$. Given $\varepsilon>0$, there exists N such that $N(\frac{1}{2}(a+1))^{N-1}<\varepsilon$, and then $|f_n(x)-f_n(a)|<\varepsilon$ for all x in $[0, \frac{1}{2}(a+1)]$ and all $n\geqslant N$ by M.V.T. So $\{f_n : n\geqslant N\}$ is equicontinuous at a. For each $i<N$, $\{f_i\}$ is equicontinuous (i.e. f_i is continuous) at a, and result follows by Exercise 6.

   Alternative: for any a in $(0, 1)$, $(f_n\,|\,[0, a])$ converges in sup metric hence any sequence in $\mathscr{F}=\{f_n\,|\,[0, a] : n\in\mathbf{N}\}\cup\{0\}$ has a convergent subsequence (in sup metric), i.e. $\mathscr{F}$ compact, and result now follows from Proposition 10.2.6.

9. No. Take $\varepsilon=\frac{1}{2}$ and any $\delta>0$, say $\delta<1$. First choose N so that

$(1-\delta)^N < \frac{1}{4}$, then choosen η with $0<\eta<\delta$ such that $(1-\eta)^N > \frac{3}{4}$, so $|1-\delta-(1-\eta)|<\delta$ but $|f_N(1-\delta)-f_N(1-\eta)|>\frac{1}{2}$.

10. For each x in $[0, 1]$ there exists $\delta(x)>0$ such that $|f(y)-f(x)|<1$ for all f in $\mathcal{F}$, all y in $B_{\delta(x)}(x)$. Take a finite subcover of $\{B_{\delta(x)}(x) : x\in[0, 1]\}$ and let Δ be the maximum $|f(x_i)|+1$ involved.
11. Similar to 10.
12. No problem: if $\mathcal{F}\subset\mathcal{C}(T, M)$ is equicontinuous then $\{a\in T : V_a \text{ bounded}\}$ is open and closed in T. In particular if T is connected then V_a is bounded either for all a in T or for no a in T.
13. Constant functions $\{n : n\in\mathbf{N}\}$ will do.
14. Let $x\in Cl(A)$ and let $\varepsilon>0$. By equicontinuity at x there exists $\delta>0$ such that $|f_n(x)-f_n(y)|<\varepsilon$ for all y in $B_\delta(x)$, all n in $\mathbf{N}$. Since $x\in Cl(A)$, there exist y in $B_\delta(x)$ such that $y\in A$, so there exists N such that $|f_m(y)-f_n(y)|<\varepsilon$ for all m, $n\geq N$. Then $|f_m(x)-f_n(x)|<3\varepsilon$ for all m, $n\geq N$. Hence x is in A.
16. Enough to show that $\mathcal{I}$ equicontinuous and closed in $\mathcal{C}(M, M)$, since M complete so $\mathcal{C}(M, M)$ complete (and M compact, so $\{f(a) : f\in\mathcal{I}\}$ relatively compact for all a in M). Equicontinuity is immediate: given $\varepsilon>0$, take $\delta=\varepsilon$. To prove $\mathcal{I}$ closed, suppose (f_n) is a sequence in $\mathcal{I}$ converging to f in $\mathcal{C}(M, M)$. Then for each x in M, $(f_n(x))$ converges to $f(x)$, so for any x, y in M, $(d(f_n(x), f_n(y)))$ converges to $d(f(x), f(y))$. But $d(f_n(x), f_n(y))=d(x, y)$ since $f_n\in\mathcal{I}$, so $d(f(x), f(y))=d(x, y)$ and by Exercise 7.3.12(a), $f\in\mathcal{I}$.

Bibliography

Ahlfors, L. V. (1966). *Complex analysis.* 2nd Edn. McGraw-Hill, New York.

Apostol, T. M. (1957). *Mathematical analysis.* Addison-Wesley Publishing Co., Reading, Mass.

Ash, R. (1965). *Information theory.* Interscience Publishers, New York.

Boas, R. P. (1969). *A primer of real functions.* The Mathematical Association of America and John Wiley & Sons, New York.

Burkill, J. C. (1970). *A first course in mathematical analysis.* Cambridge University Press, Cambridge.

Choquet, G. (1966). *Topology.* Academic Press, New York.

Copson, E. T. (1968). *Metric spaces.* Cambridge University Press, Cambridge.

Dieudonné, J. (1960). *Foundations of modern analysis.* Academic Press, New York.

Ferrar, W. L. (1938). *A textbook of convergence.* Clarendon Press, Oxford.

Herstein, I. N. (1964). *Topics in algebra.* Blaisdell Publishing Co., Waltham, Mass.

Hewitt, E. (1948). 'Rings of real-valued continuous functions I', *Trans. Amer. Math. Soc.* **64** 45–99.

——, (1960). 'The rôle of compactness in analysis', *Amer. Math. Monthly*, **67** 499–516.

Kelley, J. L. (1955). *General topology.* Van Nostrand, New York.

Kolmogorov, A. N. and Fomin, S. V. (1957). *Elements of the theory of functions and functional analysis,* Vol. 1, *Metric and normed spaces,* Graylock Press, Rochester, N.Y.

Landau, E. (1951). *Foundations of analysis.* Chelsea, New York.

Mendelson, B. (1968). *Introduction to topology.* Allyn & Bacon, Boston, Mass.

Mendelson, E. (1964). *Introduction to mathematical logic.* Van Nostrand, New York.

Patterson, E. M. (1956). *Topology.* Oliver and Boyd, Edinburgh.

Pitts, C. G. C. (1972). *Introduction to metric spaces.* Oliver and Boyd, Edinburgh.

Simmons, G. F. (1963). *Introduction to topology and modern analysis.* McGraw-Hill, New York.

Spivak, M. (1973). *Calculus.* Benjamin World Student series Edn., New York.

Bibliography

AHLFORS, L. V. (1966). *Complex analysis*. 2nd Edn. McGraw-Hill, New York.

APOSTOL, T. M. (1957). *Mathematical analysis*. Addison-Wesley Publishing Co., Reading, Mass.

ASH, R. (1965). *Information theory*. Interscience Publishers, New York.

BOAS, R. P. (1960). *A primer of real functions*. The Mathematical Association of America and John Wiley & Sons, New York.

BURKILL, J. C. (1970). *A first course in mathematical analysis*. Cambridge University Press, Cambridge.

CHOQUET, G. (1966). *Topology*. Academic Press, New York.

COPSON, E. T. (1968). *Metric spaces*. Cambridge University Press, Cambridge.

DIEUDONNÉ, J. (1960). *Foundations of modern analysis*. Academic Press, New York.

FERRAR, W. L. (1938). *A textbook of convergence*. Clarendon Press, Oxford.

HERSTEIN, I. N. (1964). *Topics in algebra*. Blaisdell Publishing Co., Waltham, Mass.

HEWITT, E. (1948). 'Rings of real-valued continuous functions I', *Trans. Amer. Math. Soc.* **64** 45–99.

——, (1960). 'The rôle of compactness in analysis', *Amer. Math. Monthly*, **67** 499–516.

KELLEY, J. L. (1955). *General topology*. Van Nostrand, New York.

KOLMOGOROV, A. N. and FOMIN, S. V. (1957). *Elements of the theory of functions and functional analysis*. Vol. 1. *Metric and normed spaces*. Graylock Press, Rochester, N.Y.

LANDAU, E. (1951). *Foundations of analysis*. Chelsea, New York.

MENDELSON, B. (1962). *Introduction to topology*. Allyn & Bacon, Boston, Mass.

MENDELSON, E. (1964). *Introduction to mathematical logic*. Van Nostrand, New York.

PATTERSON, E. M. (1956). *Topology*. Oliver and Boyd, Edinburgh.

PITTS, C. G. C. (1972). *Introduction to metric spaces*. Oliver and Boyd, Edinburgh.

SIMMONS, G. F. (1963). *Introduction to topology and modern analysis*. McGraw-Hill, New York.

SPIVAK, M. (1973). *Calculus*. Benjamin World Student series Edn., New York.

Index

Bold type indicates principal references (for example, definitions).

Ahlfors, L. V., 93, 177
Alexandroff compactification, 92
Annulus, 100
Apostol, T., 1, 121, 177
Archimedes' axiom, **4,** 9, 153
Arzelà–Ascoli theorem, 139, 141, **143,** 147
Ash, R., 23, 177

Baire's theorem, **135,** 138
Banach's theorem, 129
Basis, 49
Betweenness, 95
Bisection argument, 90, 140
Boas, R. P., 135, 177
Boundary, **64,** 71, 105
Bounded:
  function, ix, **30,** 75
  sequence, 9, 110
  set, **29,** 82
  totally, 111
Burkill, J. C., x, 177

Cantor's theorem, 134
Cartesian product, xi
Cauchy:
  convergence criterion, 9, 119
  inequality, 22
  –Schwarz–Bunyakovsky inequality, 29, 146
  sequence, **9, 109,** 123, 136, 140, **149,** 154
Choquet, G., 12, 119, 177
Closed:
  interval, xiii
  set, 59
Closure, 61
Coarser, 47
Compact, 75–77, **79,** 110, 112, 143
Compactification, 92
Complement, xi
Complete, **123,** 155, 158
Completeness axiom, **3,** 9, 81, 96, 123, 154
Completion, 155

Complex numbers, xii, 24
Component, 103
Composition, xii
Connected, 93, **94,** 102
Constant map, xii
Continuous, **13,** 19, **20, 22, 33, 36, 45,** 120
Continuous from the left (right), 15
Contraction, 130
Converge, **6, 72,** 108
Convex:
  function, **18,** 161
  set, 100
Copson, E. T., 32, 145, 177
Countable, xii
Countable basis, **50,** 71, 147
Cover, 77

De Morgan's laws, xi
Dense, **63,** 135, 154, 158
Diagonal subsequence, 141, 171
Diameter, **30,** 134
Dieudonné, J., 132, 145, 177
Dini's theorem, 122
Direct image, xi
Discrete:
  metric, 23
  space, **46,** 103
Domain, xi

Empty set, xi
ε–net, **110,** 139, 143
Equicontinuous, 142
Equivalence relation, xii, 67, 103
Equivalent metrics, 39, 88
Euclidean:
  distance, 20
  space, **22,** 86, 124
Everywhere dense, **63,** 135, 154, 158
Extension, xii

Ferrar, W. L., 11, 120, 177
Finer, 47
Fixed point, 74, 106, **129**
Frontier, 64

Function xi, 19
Function space, 26
Functor, 48, 104

Generalization, ix
Global, **77,** 87, 115, 119
Graph, **xi,** 58, 74, 92, 97, 161
Greatest lower bound, 3

Half-open interval, xiii
Hausdorff space, **73,** 82, 88, 92
Heine–Borel theorem, 79, 86
Herstein, I. N., 150, 177
Hewitt, E., vi, 78, 90, 177
Hilbert space, **30,** 125
Homeomorphism:
  of metric spaces, 41
  of topological spaces, **56,** 104

Identification map, **68,** 71
Identity map, xii
Iff, xi
Image, xi
Indexing set, xi
Indiscrete topology, 47
Induced topology, 51
Inequalities, 1, 5, 150
Infimum (inf), 3
Injective, xii
Integral equation, 132, 146
Intermediate value property (I.V.P.), **13,** 15, 93, 98, 106
Interior, **64,** 134
Interval, xiii, 79, 95
Inverse:
  function, xii, 88
  image, xi, 36
Irrational, 4
Isometry, **41,** 113, 148, 155, 157

Kelley, J. L., 23, 73, 177
Kolmogorov, A. N., and Fomin, S. V., 32, 132, 146, 177

L^1, 28, 126
L^2, 29, 146
ℓ^2, 30
Lacey, A. A., vi
Landau, E., 149, 177
Least upper bound, 2, 81
Lebesgue number, 111
Left-hand limit, 12
Limit:
  of a function, 11
  of a sequence, 7
  point, **61,** 109
Lipschitz:
  condition, 129
  equivalence, **41,** 123
  equivalent metrics, **39,** 91
Local, 77, 87, 132
Locally:
  compact, 92
  connected, 105
  constant, 105
Lower:
  bound, 2
  semi-continuous, 92

Map, xi, 19
Mendelson, B., 101, 177
Mendelson, E., 149, 177
Metric space, 21
Metrizable, 46
Möbius band, 65
Monotonic, 8
Moving bump, **118,** 140, 142

Natural numbers, xii
Net, **110,** 139, 143
Normal, **74,** 90
Nowhere dense, **64,** 134
Null sequence, 150

One–one correspondence, xii
One–point compactification, 92
Onto, xii
Open:
  ball, 33
  cover, 77
  interval, xiii
  set (in a metric space), 34
  set (in a topological space), 45
  spherical neighbourhood, 33
Order, **150,** 152
Ordered field, 1, 150

Partition, 94
Path, 100
Path-component, 105
Path-connected, 93, **100,** 102
Patterson, E. M., 177
Picard's theorem, 132
Pitts, C. G. C., 32, 132, 135, 177
Pointwise convergence, 115
Polynomial, 15,
Positive integers, xii
Powell, S., vi

Precompact, **111,** 141
Product space:
  metric, **26,** 136
  topological, **54,** 84, 98
Projection, **53,** 55, 92, 98

Quotient:
  space, 65, **68,** 73
  topology, 68

Rational numbers, xii
Real:
  numbers, xii, 1, 149
  line, xii
Regular, **74,** 90
Relatively compact, **83,** 145
Relative topology, 51
Restriction, xii
Right-hand limit, 11
Rotation-invariant, 44

Second countable, **50,** 71, 147
Sequence, 5
Sequential compactness, **108,** 112
Series, 10
Set, xi
Simmons, G. F., 86, 177
Space:
  metric, 21
  topological, 45
Spivak, M. x, 1, 177
Step function, 144
Strain, P., vi
Sub-basis, 50
Subcover, 77
Subsequence, 108, 124, 140
Subset, xi
Subspace:
  metric, 24
  topological, 51
Successive approximations, 6, 132
Supremum (sup), 3
Sup metric, **28, 31,** 119, 126
Synthetic basis, 49

Topologically equivalent metrics, 39
Topology, 45
Topological:
  invariant, 59
  property, 59
  space, 45
Torus, 65
Totally:
  bounded, **111,** 139, 143
  disconnected, 103
Translation-invariant, 44
Triangle inequality, 5, **21,** 153
Tychonoff:
  condition, 73
  theorem, 86

Uniform:
  continuity, **87,** 91, 130, 136, 158
  convergence, 115
  equivalence, 88
  metric, 28
Uniformly:
  Cauchy, 119
  equicontinuous, **142,** 144, 148
  equivalent metrics, **88,** 91, 136, 147
Union, xi
Uniqueness (of limit), 8, 72, 73
Upper bound, 2

Weaker, 47
Weak topology, 51, 56
Weyl, H., 79

Zariski topology, **48,** 74, 89

One of the ways in which topology has influenced other branches of mathematics in the past few decades is by putting the study of continuity and convergence into a general setting. This book introduces metric and topological spaces by describing some of that influence. The aim is to move gradually from familiar real analysis to abstract topological spaces. After establishing the language of metric and topological spaces with continuity as the motivating concept, the discussion develops to cover compactness, connectedness, and completeness. The book is aimed primarily at the second-year mathematics student, and numerous exercises are included.

The author is a Fellow of New College and Lecturer in Mathematics in the University of Oxford. He has previously taught at the Massachusetts Institute of Technology and at Manchester University.

FROM REVIEWS

This is a well written and to be recommended text . . . The book contains an excellent collection of exercises together with a guide to help the reader through these problems. *Bulletin of the IMA*

. . . a carefully written introduction . . . a good, informative, well organized book . . .
Jahresbericht der Deutschen Mathematiker-Vereinigung

OTHER TEXTBOOKS PUBLISHED BY OXFORD UNIVERSITY PRESS

Elementary geometry
John Roe

Groups and geometry
Peter M. Neumann, Gabrielle A. Stoy, and Edward C. Thompson

Optimal control and the calculus of variations
Enid R. Pinch

Probability: an introduction
Geoffrey Grimmett and Dominic Welsh

Introduction to complex analysis
H. A. Priestley

ISBN 0-19-853161-3

9 780198 531616

OXFORD UNIVERSITY PRESS